University of Plymouth

PEARL **https://pearl.plymouth.ac.uk**

04 University of Plymouth Research Theses 01 Research Theses Main Collection

2000

A MOLECULAR APPROACH TO CALANUS (COPEPODA:CALANOIDA) DEVELOPMENT AND SYSTEMATICS

Lindeque, Penelope Kate

http://hdl.handle.net/10026.1/2641

http://dx.doi.org/10.24382/4825
University of Plymouth

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

A MOLECULAR APPROACH TO *CALANUS* (COPEPODA: CALANOIDA) DEVELOPMENT AND SYSTEMATICS

by

Penelope Kate Lindeque

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

Doctor of Philosophy

Department of Biological Sciences
Faculty of Science

In collaboration with
Plymouth Marine Laboratory

January 2000

900432708 5

UNIVERSITY OF PLYMOUTH
Item No. 900 4327085
Date 14 SEP 2000
Class No. T 595.34 LIN
Contl. No. X 70412 3559
LIBRARY SERVICES

REFERENCE ONLY

LIBRARY STORE

For Petrus

And

In memory of my parents

I believe love survives death…and your passing, although sad and unacceptable, has been a reminder to us all to make the most of every moment. You are forever present.

"Success consists of getting up just one more time than you fall."

Oliver Goldsmith

A MOLECULAR APPROACH TO *CALANUS* (COPEPODA: CALANOIDA) DEVELOPMENT AND SYSTEMATICS

Penelope Kate Lindeque

ABSTRACT

Production and recruitment measurements in marine copepods of the genus *Calanus* have been addressed via the study of genes involved in early embryogenesis. The first sequence from a *Calanus helgolandicus* (*C. helgolandicus*) developmental gene (*Cal-Antp*) has been cloned by screening a *C. helgolandicus* genomic library with a homologous *Calanus* homeobox probe. Sequencing of an isolated and sub-cloned fragment of this gene, plus further analysis by Inverse Polymerase Chain Reaction (IVPCR), has shown it to be homologous with other *Antennapedia* homeobox genes. The temporal expression of *Cal-Antp* was analysed through its messenger RNA (mRNA) complement by Reverse Transcription Polymerase Chain Reaction (RT-PCR). The gene was expressed in tissue taken from eggs over 18 hours old, and in nauplii and copepodite stages, but no expression was detected in eggs less than 18 hours old or adult tissue. Three further homeobox-containing genes have been identified and analysed through their expression in *C. helgolandicus* eggs. Two of these are *caudal* homologues, and the third is homologous to the *Antennapedia* class of genes. The *C. helgolandicus* developmental gene sequence data provides a means of developing probes to monitor the temporal expression of such genes and their responses to environmental influence. The applicability of such probes to the investigation of key production and recruitment processes, including egg viability measurement, is discussed.

A relatively simple and cost effective method has been developed to identify the four *Calanus* species common to the North Atlantic. This system involves the PCR amplification of a region of the mitochondrial rRNA gene without prior purification of the DNA, followed by Restriction Fragment Length Polymorphism (RFLP) analysis of the amplified product. The versatility of the method is demonstrated by the unambiguous identification to species of any life stage, from egg to adult, and of any individual body parts. The molecular identification technique has for the first time shown the unexpected presence of three different *Calanus* species in Lurefjorden, Norway and has proved to be consistently accurate for all individuals tested including geographically distinct conspecific populations.

LIST OF CONTENTS

List of Figures

List of Tables

ABBREVIATIONS

AFLP	Amplified fragment length polymorphism
APS	Ammonium persulphate
bp	Base pair
B + W	Bind and wash
BSA	Bovine serum albumin
CI, CII, CIII, CIV, CV	Copepodite stages I, II, III, IV, V
C. helgolandicus	*Calanus helgolandicus*
C. finmarchicus	*Calanus finmarchicus*
C. glacialis	*Calanus glacialis*
C. hyperboreous	*Calanus hyperboreous*
°C	Degrees centrigrade
cDNA	complementary DNA
CIAP	Calf intestinal alkaline phosphatase
CSPD	Disodium 4-chloro-3-(methoxyspiro{1,2-dioxetane-3-2'-(5'-chloro)tricyclo[3.3.1.1 ^{3,7}] decan}-4-yl)phenyl phosphate
DAPI	4',6-diamidino-2-phenylindole
dATP	Deoxyadenosine 5'triphosphate
dCTP	Deoxycytidine 5'triphosphate
dGTP	Deoxyguanosine 5'triphosphate
dUTP	Deoxyuridine 5'triphosphate
dTTP	Deoxythymidine 5'triphosphate
dNTP	Deoxynucleotide 5'triphosphate
ddNTP	Dideoxynucleotide 5'triphosphate
DEPC	Diethylpyrocarbonate
DIG	Digoxigenin
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
Dnase	Deoxyribonuclease
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetracetic acid
EMBL	European molecular biology laboratory
FLP	Fragment length polymorphism
gcg	Genetics computer group
g, mg, µg, ng	Gram, milligram, microgram, nanogram
h	Hour
IMS	Industrial methylated spirits
IPTG	Isopropyl thiogalactoside
IVPCR	Inverse PCR
Kb	Kilobase pair
Km, cm, mm, µm	Kilometre, centimetre, millimetre, micrometre
λ, λ phage	λ bacteriophage
LB	Luria-Bertani Broth
L, mL, µL	Litre, millilitre, microlitre
Lids	Lithium-dodecyl-sulphate
ln	Natural log
Ltd	Limited
M, mM, µm	Molar, millimolar, micromolar

min	Minute
MOPS	3-[*N*-morpholino] propanesulphonic acid
mRNA	Messenger RNA
mt DNA	Mitochondrial DNA
NI, NII, NIII, NIV, NV, NVI	Naupli stages I to VI
-ve	Negative
OD	Optical density
PCR	Polymerase chain reaction
pfu mL^{-1}	Plaque forming units per mL
+ve	Positive
RAPD	Random amplified polymorphic DNA
rATP	Rat adenosine 5'triphosphate
RCF	Relative centrifugal force
rDNA	Ribosomal DNA
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
Rnase	Ribonuclease
rpm	Revolutions per minute
RT-PCR	Reverse transcription-PCR
Sec	Second
SDS	Sodium dodecyl sulphate
Spi	Sensitive to P2 inhibition
TAE	Tris-acetate-EDTA buffer
TBE	Tris-boric-EDTA buffer
TE	Tris-EDTA buffer
TEMED	N,N.N',N'-tetramethylethylenediamine
Tm	Melting temperature
Tris	Tris[hydroxymethyl]-aminomethane
Tween 20	Polyoxyethylenesorbitan monolaurate
U	Units
µF	Microfarads
UV	Ultrviolet
V	Volts
w/v	Weight per volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Acknowledgements

Firstly I would like to thank my supervisors Gary Smerdon, Roger Harris, and Malcolm Jones for making this PhD project possible, and for their help and guidance over the past four years. I would especially like to thank Gary for his exhaustive training in molecular techniques and for his patience with my never ending questions. Your help and friendship have been much valued.

I would like to acknowledge the help of members of the molecular biology laboratory at Plymouth University for their advice and help with caesium chloride and sucrose gradient centrifugation. Thanks also go to the work experience students, Hannah Gibson and Joanne Shepherd for their interest and help in the species identification technique.

Huge thanks also, to all past and present members of the molecular biology laboratory (G17) for all your support, encouragement and numerous cups of coffee and pints of beer. To the inhabitants of room 109, for chats, support and for the haven of somewhere quiet to work, thank you.

A special thank you goes to Alan Bell, Petrus Lindeque and Xabier Irigoien for proof reading various chapters of this thesis. Thank you also to big brother, for helping with the cutting out and sticking of the pictures!

I would also like to thank all friends and members of PML involved in squash playing, mountain biking and sailing, for making Plymouth such a fun place to work, and to friends at home, for keeping me sane over the past four years.

The biggest and most heartfelt thank you of all has to go to my husband Petrus. Without whose constant love, faith, patience and encouragement none of this would have been possible. "I promise I'll be finishing soon...."

Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award.

This study was financed with the aid of a studentship from the Natural Environmental Research Council special topic PRIME (Plankton Reactivity in the Marine Environment), and was carried out in collaboration with Plymouth Marine Laboratory.

A programme of advanced study was undertaken, which included extensive training in a wide range of molecular biology techniques including plasmid cloning, restriction mapping, Southern blotting, Polymerase Chain Reaction, bacteriophage cloning, genomic library construction, mRNA analysis (using reverse transcriptase PCR), DNA sequence analysis and computer-based data analysis of nucleic acid sequences. Training was also obtained in the use of information technology software programmes.

Relevant scientific seminars and conferences were regularly attended at which work was often presented, and a paper prepared for publication.

Publications:

1) Lindeque, P. K., Harris, R. P., Jones, M. B., and Smerdon, G. R., 1999. Simple molecular method to distinguish the identity of *Calanus* species (Copepoda: Calanoida) at any developmental stage. Mar. Biol., **133**, 91-96.

Meeting and Conferences attended:

1) PRIME mid-term symposium. University of Warwick, April 1997.
Oral presentation: 'Copepod production in the marine environment – A molecular approach.'

2) Research seminar series. University of Plymouth, September 1997.
Oral presentation: 'Copepod production in the marine environment – A molecular approach.'

3) Second annual meeting of TASC (TransAtlantic Study of *Calanus*). Bremerhaven, Germany, January, 1998.
Poster presentation: '*Calanus* species identification: a simple molecular method to distinguish both adult and pre-adult stages.'

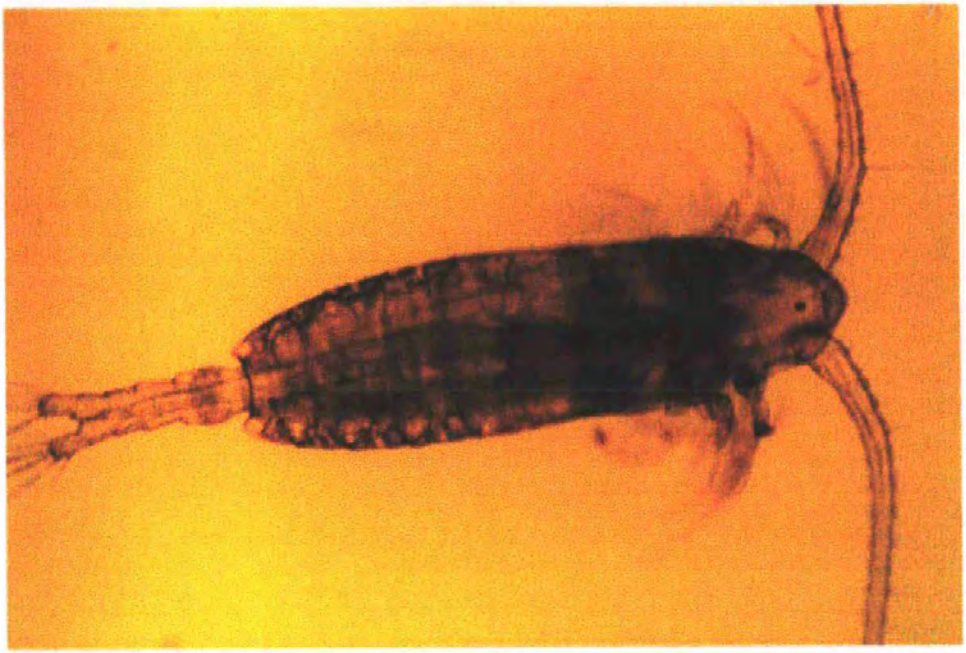
4) PRIME final symposium. University of Bangor, April, 1998.
Poster presentation: '*Calanus* production and systematics – A molecular approach.'

5) Third annual meeting of TASC. Villefrance, France, January, 1999.
Oral presentation: '*Calanus*: development and application of a simple molecular identification technique.'

6) Research seminars, final year presentation. University of Plymouth, May, 1999.
Oral presentation: 'A molecular approach to *Calanus* (Copepoda: Calanoida) development and systematics.'

7) ICES symposium: Population dynamics of *Calanus* in the North Atlantic. Results from the Trans-Atlantic study of *Calanus finmarchicus*. Tromsø, Norway, August, 1999.
Oral presentation: 'A molecular approach to *Calanus helgolandicus* development.'

Signed: P.R. Lindeque
Date: 21·01·00



Calanus helgolandicus – adult female

(actual size 2 – 4 mm)

CHAPTER 1

Introduction

CHAPTER 1

Introduction

1.1 Copepods and the zooplankton community

1.1.1 Zooplankton

1.1.2 The contribution of copepods to secondary production

1.1.2.1 Calanus

1.1.3 Methods of measuring secondary production

1.1.4 *Calanus* egg viability and recruitment

1.2 New techniques and technologies

1.2.1 Current techniques available

1.2.2 Problems associated with using molecular techniques in marine biology

1.3 *Calanus* development

1.3.1 Overview

1.3.2 *Calanus* developmental genes

1.3.2.1 Model systems

1.3.2.2 DNA binding proteins - The homeobox

1.3.2.3 Application of model systems to obtaining Calanus *developmental genes*

1.3.3 Potential use of *Calanus* developmental genes

1.4 *Calanus* systematics

1.4.1 Current molecular applications

1.4.2 Potential molecular applications

1.5 Aims and objectives

1.1 Copepods and the zooplankton community

1.1.1 Zooplankton

Fundamental to the understanding of marine ecosystems is knowledge of the pelagic food web. Central and important links in this food web are played by zooplankton. Zooplankton are the animal population of plankton, plankton being characterised by the fact that they drift and float with water movement within the marine environment (Marshall and Orr, 1972). It is usually thought that zooplankton have little or no power of independent horizontal migration, although they are able to migrate vertically and hence inhabit all layers of the ocean down to the greatest depth sampled (Marshall and Orr, 1972; Raymont, 1983).

Of all the marine zooplankton the class of crustacean known as Copepoda are the dominant constituent of the plankton in every sea and ocean, usually compromising at least 70% of the plankton fauna and representing up to 90% of the biomass in shelf areas (Jaschnov, 1970; Raymont, 1983).

1.1.2 The contribution of copepods to secondary production

With the dominance of copepods in the zooplankton community it is not surprising that they play an important role as secondary producers in the pelagic food web. For a considerable period of time the link between diatoms as primary producers, and copepods as secondary producers, has been considered a key element in the classic pelagic food chain (Kleppel, 1993). A new understanding, however, of the relationships between copepods and the food environment has been emerging over the last two decades. This has, for example, emphasised the association between food nutritional composition and the nutritional needs of the feeders, rather than simply the carbon balance between them (Kiorbøe *et al.*, 1985; Jónasdóttir, 1994; Ianora *et al.*, 1995; Ban *et al.*, 1997; Jónasdóttir *et al.*, 1998), and has shown that energy flows through complex food webs rather than along

simple food chains. Copepods form an important route for the transport of carbon through these food webs by processing photosynthetically produced organic matter and passing it up the food web to higher trophic levels, and down through the water column in the form of faecal pellets and excretion products. The production of copepod eggs and nauplii is a major route for particulate organic matter transfer from phytoplankton to fish larvae and small fish species (Runge, 1988; Runge and Lafontaine, 1996).

1.1.2.1 Calanus

Copepods of the genus *Calanus* form a significant proportion of the zooplankton biomass in the North Atlantic (Jaschnov, 1970). *Calanus* play an important role as secondary producers in the marine food web, grazing extensively on phytoplankton and representing significant prey species for larvae, juvenile and adults of commercial fish (Runge, 1988; Dickson and Brander, 1993; Conover *et al.*, 1995; Runge and Lafontaine, 1996; Orlova, 1999). The four *Calanus* species in the North Atlantic, *Calanus helgolandicus*, *Calanus finmarchicus*, *Calanus glacialis* and *Calanus hyperboreous*, show a wide area of distribution, and although having distinct hydrographic affinities are sympatric in many areas with overlapping reproductive periods. *C. helgolandicus* is a warm-temperate water species, found in southern areas of the North Atlantic, the Celtic Sea, North Sea and in coastal waters south of England (Fleminger and Hulsemann, 1977; Planque and Fromentin, 1996). *C. finmarchicus*, although a more cold-temperate water species, co-occurs extensively with *C. helgolandicus* in many areas of the North Atlantic, and in shelf Seas around the United Kingdom (Williams and Conway, 1980). The distributions, interactions and different life strategies of these two *Calanus* species are complex, they show converse vertical distributions in areas where they co-exist (Williams and Conway, 1980) and their biogeographical boundaries are modified by the North Atlantic Oscillation (Fromentin and Planque, 1996). *C. glacialis* and *C. hyperboreous* are

Arctic water species found at more northern latitudes. They are sympatric with *C. finmarchicus* in many areas including the region of the polar front (Frost, 1974). In the eastern North Atlantic the former two species also co-occur in the region of the Gulf of St. Lawrence, Hudson and Davis Straits (Bucklin *et al.*, 1995).

1.1.3 Methods of measuring secondary production

Secondary production can be defined as the total production of organic matter by heterotrophic organisms over a given period of time (Kimmerer, 1987). Pelagic secondary production encompasses a wide range of organisms from planktonic bacteria to fish, and thus no single technique exists for its measurement. No equivalent *in situ* method such as the ^{14}C technique used to measure primary productivity exists for secondary production measurements, and attempts to predict production levels from primary production measurements have proved largely unsuccessful.

With zooplankton constituting a consistently significant proportion of heterotrophic biomass there have been several approaches applied to the measurement of zooplankton secondary production. A holistic method of measurement known as cohort analysis involves following changes over time in numbers and growth, under natural conditions, in a single cohort of a population (Huntley and Lopez, 1992). In practical terms there are many limitations to cohort analysis. For example, most natural zooplanktonic populations are not clearly composed of true cohorts. Thus, sampling a true cohort would require that samples be taken with adequate resolution in time and space (Huntley and Lopez, 1992). Alternatively, secondary production can be viewed as the sum of the whole instantaneous rate of biomass production, realised by the different zooplankton species (Runge and Roff, 1999). Such measurements rely on estimations of the quantity of living tissue present (biomass) and the rate of synthesis of new tissue (growth rate). Whilst instantaneous

estimations of biomass are feasible, growth rate determinations can prove much more complex.

The search for a suitable technique with which to estimate zooplankton growth rates is on-going. Many measurements have involved 'the physiological method', or derivatives thereof, whereby growth is calculated as a function of assimilation efficiency, clearance rate, food concentration and ingestion rate (Huntley and Lopez, 1992). The impetus for the physiological method is the assumption that all physiological rates that influence individual growth respond in a quantitative and predictable manner to a variety of environmental factors. In reality, the physiological rates of individuals do not respond instantaneously to their immediate environment and, together with experimental error, this can mean physiological methods can yield untenable results (Huntley and Lopez, 1992). Several empirically or allometrically derived approaches to the estimation of planktonic growth, which can give a basic approximation of the growth of zooplankton, have also been proposed (Runge and Roff, 1999). The search for a suitable growth rate measurement has also included biochemical techniques, such as measurements of RNA and DNA (Sutcliffe, 1965; Dagg and Littlepage, 1972; Saiz *et al.*, 1998). The potential use of nucleic acids to estimate the rates of growth of marine organisms in nature relies on the assumption that the amount of DNA per cell is constant while RNA fluctuates in proportion to metabolic (including growth) rate. There is often a positive correlation between RNA content and growth rate, but in several studies the relationship has been shown to lack sufficient specificity to predict reliably growth rates (Ota and Landry, 1984; Wagner *et al.*, 1999).

Sapienza and Mague (1979) demonstrated a positive correlation between growth rate and DNA polymerase activity in *Artemia,* thereby introducing an alternative approach to measuring growth rates using enzymatic indices. Further studies of enzymatic indices have shown a reasonable relationship between the enzyme aspartate transcarbamylase

(ATCase) and growth rates (Bergeron and Alayse-Danet, 1981; Bergeron, 1982; Bergeron 1995). However, in many cases the correlation between enzyme activity and growth rates has been uncertain and unreliable (Hérnandez-Léon *et al.*, 1995; Biegala *et al.*, 1999).

Currently, a number of assessments of copepod growth rate rely on measurements of *in situ* egg production rates in adult female copepods. This method is a practical simplification of the physiological method. Kiørboe and Johansen (1986) and Berggreen *et al.* (1988) suggested that production could be measured by multiplying the specific egg production rate of female copepods by the total biomass, assuming that all stage specific instantaneous growth rates are equal and can be estimated from the egg production rate of females. It has been shown for several species of marine copepod that, since adult females have ceased to grow, the rate of specific egg production for adult female copepods closely approximates to the growth rate of earlier developmental stages (Corkett and McLaren, 1978; Sekiguchi *et al.*, 1980; Berggreen *et al.*, 1988). The production of eggs by adult female *Calanus* gives an indication of the functional reproductive responses of copepods to environmental hydrographic and phytoplankton food conditions (Kiørboe *et al.*, 1988; Ianora and Buttino, 1990; McKinnon and Thorrold, 1993; Laabir *et al.*, 1995).

The basic procedure for estimating the egg production rate of *Calanus* involves incubating recently captured females in seawater containing a natural assemblage of phytoplankton (Runge, 1985; Beckman and Peterson, 1986). After an incubation time of normally 24 hours, at *in situ* temperature, the eggs produced per female are counted.

Measurements of *in situ* egg production rates in adult female copepods have become a widely used and accepted technique to measure growth rates. Such measurements have also been used for the estimation of factors controlling the dynamics of copepod populations, such as mortality and recruitment. However, this relies on the assumption that all eggs produced by adult females will hatch and lead to a new generation of nauplii (Marshall and Orr, 1952). The principle has now been reconsidered (Ianora and

Poulet, 1993; Poulet *et al.*, 1994) since it reflects neither the fate of secondary production nor the fertility of the adults. The fate of the eggs in terms of viability is a major consideration since it has been demonstrated that up to 100% of eggs produced can be non-viable depending on the conditions encountered in the field or maintained in the laboratory. For accurate measurements of copepod population dynamics, it is therefore imperative that egg production measurements are accompanied by measurements of egg hatching success.

1.1.4 *Calanus* egg viability and recruitment

The numerous egg production studies of zooplankton that have been performed (Marshall and Orr, 1972, onwards) normally focus on fecundity *per se*. However, two equally important aspects of the reproductive biology of copepods, including the capacity of the female to produce viable eggs and the survival of offspring, have not, until recently, been as thoroughly investigated. The study of copepod egg viability is essential, since variability of copepod recruitment may depend largely on the rate of production of viable eggs rather than fecundity *per se* (Ianora and Poulet, 1993; Jónasdóttir, 1994; Poulet, 1994, Laabir *et al.*, 1995, Miralto *et al.*, 1999). The egg viability-recruitment relationship is necessary to determine optimal copepod survival, because this type of relationship defines the population's ability to replenish itself (Ianora and Poulet, 1993; Poulet *et al.*, 1994).

Estimates of recruitment using copepod egg production rates may sometimes be seriously overestimated, considering that the hatching success of newly spawned eggs varies between 0 and 100% (Poulet *et al.*, 1994, 1995; Ianora *et al.*, 1995). Recent studies have identified several potential explanations for reduced viability in copepod eggs, including female nutrition, food quantity, quality, composition, size and distribution (Ambler, 1985; Ianora *et al.*, 1992; Jónasdóttir, 1994; Guisande and Harris, 1995; Jónasdóttir and Kiørboe, 1996; Pond *et al.*, 1996; Miralto *et al.*, 1998), and the production of inhibitory compounds by diatoms (Poulet *et al.*, 1994; Miralto *et al.*, 1995; Ianora *et al.*,

1995; Chaudron *et al.*, 1996; Ianora *et al*, 1996; Uye, 1996, Ban *et al.*, 1997; Miralto *et al.*, 1999). Other extrinsic factors that can affect egg viability include temperature (Beckman and Peterson, 1986; Kiørboe *et al.*, 1988; Kleppel, 1993), salinity (Trujillo-Ortiz, 1996), anoxia (Ambler, 1985) and physical environmental conditions (Peterson, 1985; Scrope-Howe and Jones, 1985; Tiselius *et al.*, 1990; McKinnon and Thorrold, 1993). Egg production and viability of marine copepods can also depend on intrinsic factors such as condition and age of females (Ianora and Buttino, 1990), endogenous diel rhythms (Stearns *et al.*, 1989; Checkley *et al.*, 1992) and sperm age and quality (Pond *et al.*, 1996).

It is therefore apparent that a measurement of egg viability is equally as important as egg production measurements to gain a full understanding of *Calanus* population dynamics. These two variables give an accurate measurement of a fraction of secondary production and recruitment in the field. Methods have been developed with which to assess egg viability, but the extra analysis increases the already labour-intensive nature of egg production experiments. In addition, there is always the potential of incubation artefacts affecting egg viability results. Currently the determination of egg viability involves the collection of eggs laid during egg production experiments, and their further incubation under controlled conditions. Hatching success rates are subsequently determined by microscopic examination.

There is a need for a new technique that is able to ascertain the viability of eggs. This technique must be less labour intensive and be suitable for application to eggs collected directly from the field, eliminating any potential laboratory artefacts.

1.2 New techniques and technologies

1.2.1 Current techniques available?

The spatial and temporal distributions of organisms in the oceans have been catalogued by the biological oceanographer for much of the 20th Century. More recent

studies have provided an understanding of how the physical and chemical environment interacts with biological processes in the pelagic system. However, there is a need for new techniques and technologies to be applied to marine ecosystems, in order to increase our understanding of interactions between physical, chemical and biological processes, and in particular of the important role played by copepods in pelagic marine food webs.

Three new tools that have become available include remote sensing, *in situ* moored instrumentation and molecular biology techniques. Application of remote sensing techniques and *in situ* instruments has provided insight into large spatial and temporal distributions of both phytoplankton and zooplankton in the upper ocean in relation to physical processes (Falkowski and LaRoche, 1991). Molecular biology techniques are widely used in the biomedical field, but the application of such techniques within marine biology is still very limited. It is possible that such techniques will provide a foundation for mechanistically understanding the biological basis for the distribution and function of oceanic organisms in the future.

1.2.2 Problems associated with using molecular techniques in marine biology

Molecular methods can provide a good understanding of many biological processes at a fundamental level. However, compared to other areas of biology, the application of molecular techniques to biological oceanography has lagged. Falkowski and LaRoche (1991) suggest three reasons for this:- 1) difficulty in applying standard molecular biology techniques to oceanic systems, where concentrations of nucleic acids and proteins may be extremely low and their sources poorly identified; 2) a small historical data base, combined with poorly characterised genetic systems; 3) lack of communication between oceanographers and molecular biologists. Initially, molecular biology techniques used in the ocean environment were subject to many complications. Early measurements of nucleic acids in the ocean were based on bulk measurements. For example, Holm Hansen

et al., (1968) attempted to use a fluorometric method based on the reaction of 3,5-diaminobenzoic acid with DNA to measure the amount of DNA in the water column. The results were poorly correlated with cholorophyll and the authors concluded that "there is a considerable quantity of living material that is high in DNA, or that DNA is associated with particulate, non-living material." Using DNA measurements involving DAPI-staining methods, Falkowski and Owens (1982) also concluded that there were varying amounts of detrital DNA present in the natural phytoplankton communities. Indeed, simply budgeting the nucleic acids in a natural water column remains a challenge.

Recent developments however, have brought about an abundance of new molecular biology techniques, including, for example, the use of gene cloning, gene probes, and DNA fingerprinting (Brown, 1995). Such techniques allow biological oceanographers to address many questions related to the phylogeny, distribution and biochemical processes of zooplankton. In particular this study is concerned with the expansion and application of molecular techniques to increase our understanding of the development and systematics of the important copepod genus *Calanus*.

1.3 *Calanus* development

1.3.1 Overview

Female *Calanus* have a single ovary located dorsal to the gut in the posterior part of the cephalosome. The ovary contains many oogonia undergoing mitotic division to become oocytes. As the oocytes move from the ovaries through the oviducts they mature. After oogenesis the eggs pass through the genital atrium being fertilised on the way by spermatozoa stored in the spermathecae. The physiological basis of *Calanus* embryogenesis is described by Grobben (1881). At the initial egg stage the male and female pronuclei fuse creating the first blastomere. The first cleavage that occurs is meridonal and gives rise to two blastomeres. Divisions continue at right angles to one

another and the product of the first rounds of cell division of the fertilised egg is a solid globular mass of 32 identically sized cells. Cell division continues and after 2 hours (at 15-20°C) a segmentation cavity becomes distinguished, this hollow ball of cells being known as the blastula. As embryogenesis proceeds the blastula becomes invaginated to form a cup- or basin- shaped early embryo in which the three primary germ layers (ectoderm, mesoderm, and endoderm) begin to be distinguished.

The physiological events of early embryogenesis for *Calanus* species have been understood for a number of years, partly because of the ease with which embryonic development can be seen through the clear membrane of the embryo. However, to date nothing of the molecular basis of *Calanus* development is understood. From the literature it is evident that there is an abundance of factors, both extrinsic and intrinsic, that affect the rate of egg production and viability. Although we have an understanding of many factors that create a change in the reproductive potential of copepods, and hence population fluctuations, at the gross phenotypic level, we are without any understanding of the underlying genotypic basis of such changes.

1.3.2 *Calanus* developmental genes

Over the past two decades there has been a revolution in the understanding of how living things create themselves, especially at the molecular genetic level. For a wide range of organisms there has been a vast increase in the understanding of the mechanisms by which a fertilised egg develops into an adult. Molecular techniques may be used to extend this knowledge of developmental mechanisms to include the ecologically important genus of copepod, *Calanus*. The acquisition of *Calanus* developmental genes is of considerable scientific interest. By beginning to understand the molecular mechanisms of *Calanus* development it will become possible to investigate the effect of environmental factors on

the expression of key developmental genes. Such genes would also provide unequivocal tools with which to estimate key production processes such as copepod egg viability.

Little is known about *Calanus* genes, and it is therefore necessary to turn to model systems for the investigation of *Calanus* developmental pathways. The system for which most genetic information is available is that of the fruit fly *Drosophila melanogaster.*

1.3.2.1 Model systems

The fundamental principle of developmental biology is that growth and differentiation result from genes being turned on and off. The basic body plan of animals is not specified all at once, but is formed as a result of a hierarchy of developmental decisions. The same genetic mechanisms work in different species from sponges to humans, controlling major pathways of development.

Insect embryos have been a popular subject of analysis among embryologists for some time. These studies have generated the fundamental concepts on which our current understanding of the genetic control network is based. Using combined genetic and molecular approaches (for review see Ingham, 1988; Lawrence and Morata, 1994) the principles of genetic control in *Drosophila* embryogenesis have been elucidated. These principles can be used as a model to explicate the genetic control of other taxa, in particular other arthropods such as marine copepods.

In *Drosophila*, coordinate genes are first expressed maternally during oogenesis, and determine the ventral-dorsal and anteroposterior axis (Nüsslein-Volhard, 1991). The ventral-dorsal axis is established by the initially uniform distribution of the *dorsal* protein, subsequently becoming distributed in the nuclei of the blastoderm in a ventral-dorsal gradient (Govind and Steward, 1991). A major determinant of the anteroposterior pattern is the product of the *bicoid* gene. *Bicoid* is transcribed in the nurse cells of the ovary and then passes into the nearest end of the egg, forming an initial asymmetry concentrated at

the anterior pole. The second, posterior system, of which the pole plasm is an integral part, deposits the mRNA of a gene called nanos in the prospective abdominal region. When the mature egg is laid it is thus both morphologically and molecularly polarised. After fertilisation the mRNAs are translated to form concentration gradients of proteins with opposite polarities acting as morphogens (ie different morphological structures are determined by different concentrations). The products of these maternal systems thus divide the embryo into several zones depending on their relative concentrations or activities. Before cellularisation one nucleus can affect the gene activity of nearby nuclei simply by producing a genetic regulatory protein, with no receptors or signal transduction mechanisms being required (Gehring, 1987). Because of the overlaps between the different protein domains, and because different concentrations of the same substance can have different effects, the maternal systems can activate a spatial pattern of zygotic gene activity which is more complex than their own (Wieschaus, 1996).

These zygotic genes, which lead to the conversion of the gradient patterns into a periodic pattern of body segments, are known as 'segmentation' genes. The segmentation genes can be divided into three classes, the gap genes, pair-rule genes, and segment polarity genes. The gap genes are so called because their mutant phenotypes consist of the elimination of particular regions of the body, thereby creating a gap in the antero-posterior pattern. There are three principle members of this class, *hunchback*, *Krüppel* and *knirps*, each affecting regions several segments wide (Ingham, 1988). The pair-rule genes are expressed in a stripy pattern with a periodicity of two segment widths. The pair-rule genes such as *hairy*, *even-skipped*, and *runt*, in various concentrations with the maternal and gap gene products then activate the segment polarity genes. In turn the combined maternal, gap, pair-rule and polarity gene product combinations activate the homeotic selector genes of the Antennapedia complex and Bithorax complex which control the character of each parasegment and thus its subsequent pathway of differentiation (Slack, 1997). The

homeotic genes are classically defined by their mutant phenotypes, which consist of the transformation of one segment type into another (Ingham, 1988). The targets of the homeotic selector genes are poorly defined, but their activation represents a transition from global regulation of body plan by the maternal and early zygotic products to differential specification of tissue types and ultimately body structures (Reid, 1990).

The *Drosophila* model of developmental genetic mechanisms is invaluable in increasing understanding of the molecular events involved in embryogenesis in a wide range of invertebrates and vertebrates. However, account must be taken of the differences between the physiological process of early development in *Drosophila* when compared to other invertebrates and vertebrates. In many insects, segments are generated sequentially as the cells of the embryo proliferate. By contrast, in *Drosophila* and other so-called long germ-band insects, the entire body plan is established simultaneously at the blastoderm stage of embryogenesis. After fertilisation a rapid period of synchronous nuclear divisions occurs without cellular cleavage. After nine divisions most of the nuclei migrate to the periphery to form the syncytial blastoderm. The embryo remains a syncitium while the nuclei undergo three further divisions. Membranes then grow down from the cortex to separate each nucleus, so forming the cellular blastoderm. By the completion of this process, the fate of each newly formed cell is already well established (for details see Gehring, 1987; Slack, 1997).

Despite apparent differences in body plan and large evolutionary separation of organisms, the patterning principles derived from studies of the fruit fly are remarkably universal. *Bicoid* and other morphogens can diffuse rapidly in the fluid mass of a fruit fly embryo, whereas in a multicellular embryo any corresponding processes would have to involve intercellular signalling (Slack, 1997). However, the developmental mechanisms in diverse organisms are still more homologous than could ever have been predicted (for reviews see Reid, 1990; Scott, 1994). The common genetic structure in developmental

pathways is clearly seen in the homeotic genes that play a crucial role in signal cascades, serving as master switches, or gene controllers. These homeotic genes encode proteins that do not become muscle or bone, but rather they bind to the DNA of other genes, helping to express or silence them and thus directing the emergence of the body plan. The homeotic genes' binding proteins represent the main group of transcriptional regulators controlling gene expression during embryogenesis (Lewin, 1997).

1.3.2.2 DNA binding proteins - The homeobox

Comparisons between the sequences of many transcription factors or DNA-binding proteins suggest that there are common types of motifs. These motifs are responsible for binding to the DNA, and include steroid receptors, zinc fingers, helix-turn-helix, and helix-loop-helix motifs. The motifs are usually quite short and comprise only a small part of the protein structure.

The products of the homeotic selector genes (and others) are transcription factors or DNA-binding proteins with a common motif. The discovery of this common motif, known as the homeobox (McGinnis *et al.*, 1984; Scott and Weiner, 1984) has been the key to developmental biology. The homeobox is a conserved DNA sequence approximately 180 base pairs in length. It encodes a highly conserved polypeptide segment of 60 amino acids, known as the homeodomain (McGinnis *et al.*, 1984). This defined domain mediates the binding to specific DNA sequences, whereby the homeotic proteins exert a gene regulatory function. Each homeotic gene contains a homeobox, however there are homeobox-containing genes at all levels of development, from maternally expressed genes such as *bicoid*, through gap genes and segmentation genes.

The homeodomain includes a helix-turn-helix motif similar to the DNA binding domain of prokaryotic transcriptional regulators (Laughon and Scott, 1984). The homeodomain folds into three α-helices (I, II, and III) creating an internal hydrophobic

core. The recognition helix inserts into the major groove of DNA and makes sequence-specific contacts. The N-terminus of the homeodomain makes contact with the minor groove and stabilises the association with the DNA and the loop between helices I and II interact with the DNA backbone (Gehring *et al.*, 1994).

The homeobox was discovered through the structural analysis of the *Antennapedia* gene in *Drosophila* (McGinnis *et al.*, 1984; Scott and Weiner, 1984). The significance of the high homology of the homeobox was demonstrated by isolating previously unknown homeotic genes from *Drosophila* with the homeobox as a probe. More importantly the homeobox has provided an entry point to isolating genes that control development in higher organisms, on the basis of their partial homology to the *Drosophila* homeobox.

The resemblance of homeotic genes in *Drosophila* and other organisms is not only structural but also positional and functional. Although information on vertebrate developmental gene functions is limited, evidence from mice and nematodes clearly supports conservation of function of such genes (Manak and Scott, 1994). In all animals studied to date there is a cluster of homeobox genes, known as HOM-C in insects, nematodes and other invertebrates, and *Hox* in mammals (McGinnis and Krumlauf, 1992) (*Hox* is often used to refer to all such clusters). All of the homeotic genes are found within these HomC/*Hox* clusters. In invertebrates, only a single *Hox* gene cluster has been found (although it is split in *Drosophila*). The common ancestor of all chordates is assumed to have had a single cluster as well. This cluster is thought to have duplicated to four clusters on different chromosomes, accompanying the increasing complexity of body plans during evolution of vertebrates (Meyer, 1998). Comparative studies of *Hox* genes in various metazoans indicate that many *Hox* genes predate the divergence of animals at an early evolutionary stage (Grenier *et al.*, 1997). Homeotic genes that specify middle body regions in insects, two of which are unique to arthropods, *Ultrabithorax* and *Abdominal-A* (Grenier *et al.*, 1997; Mouchel-Vielh *et al.*, 1998), originated before the divergence of the

insect and crustacean lineages (Averof and Akam, 1993). Indeed, such studies indicate that crustaceans and insects share almost identical complements of *Hox* genes. Diversity of *Hox* genes within arthropods, insects and vertebrates has arisen primarily through regulatory evolution. The creative potential of regulatory evolution lies in the hierarchical and combinatorial nature of the regulatory networks that guide the organisation of body plans and the morphogenesis of body parts (for review see Caroll, 1995).

The *Hox* genes within the HomC/*Hox* clusters are arranged along the chromosome as they are expressed along the anterioposterior axis of the embryo (Gaunt *et al.*, 1988; Duboule and Dollé, 1989, Graham *et al.*, 1989). To date there are no clear exceptions to this rule (Duboule, 1998). The discovery of this colinearity in different organisms has meant that the connection between structure and function of the HOM-C complex for *Drosophila* is likely to be conserved in all other HOM-C/*Hox* complexes.

The DNA-binding homeodomain is now viewed as a hallmark of transcription factors which control the development architecture in organisms as diverse as yeasts, plants, insects and mammals (Manak and Scott, 1994; Scott, 1994). The surprising conservation of the regulators stands in stark contrast to the diversity of animal form, and has posed some interesting questions regarding evolution (Caroll, 1995; Holland and Garcia-Fernàndez, 1996; Abouheif, 1997; Averof and Patel, 1997; Meyer 1998; Rosa *et al.*, 1999; and others). *Hox* gene distribution has provided an important new tool for addressing long-standing issues in metazoan phylogeny. The homeodomain proteins, particularly the HOM/*Hox* group, remain the most dramatic example of retention of both protein structure and function during the evolution of developmental processes (Manak and Scott, 1994).

1.3.2.3 Application of model systems to obtaining Calanus developmental genes

The model of *Drosophila* developmental mechanisms and the highly conserved sequence motif of the homeobox provide ideal tools for the acquisition of *Calanus* developmental genes. Given that many homologous developmental genes have already been cloned in other organisms, and that many of the target genes share the common sequence motif of the homeobox, there are several approaches that may be used to obtain *Calanus* developmental genes. Putative early developmental genes may be acquired by probing a genomic library, constructed from the organism under study. The library may be probed with heterologous probes designed from specific early developmental genes from other animals, for example, *Drosophila.* Such genes may be chosen with knowledge of their roles in the model *Drosophila* developmental system. Alternatively, analysis of conserved regions within homeoboxes from a range of different organisms, may be used to design degenerate oligonucleotide primers, and used to amplify the equivalent portion of gene from *Calanus.* This small but specific region can then be used as a homologous probe to screen the genomic library with an intermediate level of specificity. It is also possible to identify genes through their mRNA compliment by targeting time-specific expression. Performing reverse transcriptase PCR on RNA extracted from eggs can target genes expressed in early embryogenesis.

1.3.3 Potential use of *Calanus* developmental genes

Embryogenesis plays a critical role in copepod recruitment, and yet the molecular basis of this developmental phase is not understood. The acquisition of developmental genes would enable an initial perception of the genotypic basis of *Calanus* developmental mechanisms, providing improved techniques for the key measurements of *Calanus* population dynamics, such as *Calanus* egg viability. Understanding the mechanisms of *Calanus* development would enable an assessment of the effect of environmental factors

on the expression of key developmental genes. A suitable *Calanus* developmental gene expressed in early embryogenesis would also allow the development of a molecular system with which to test egg viability. The potential viability of a relatively large number of eggs could be tested simultaneously to determine whether they are undergoing normal development, or whether they are non-viable with no developmental pathway being followed. Such a system would greatly increase our understanding of the factors affecting egg hatching success and hence copepod recruitment. In addition, it would also permit the analysis of eggs collected directly from the field, thus eliminating the potential problem of laboratory artefacts.

1.4 *Calanus* systematics

1.4.1 Current molecular applications

In the field of oceanography, molecular biology has had its greatest impact so far in the study of phylogeny and taxonomy of marine organisms. The basic technique involves determination of the similarity of nucleic acid sequences between organisms. Phylogenetic relationships can then be inferred from statistical analysis of these sequences. Most sequencing is performed on specific genes, either of nuclear or mitochondrial origin. Perhaps the most commonly used are ribosomal RNA (rRNA) genes as these are universally distributed, and are functionally and evolutionary homologous (but not identical) in all organisms. They are readily identifiable by their sizes, constitute a significant part of the cellular mass, and are easy to isolate from many types of organisms.

The molecular systematics of copepods, including *Calanus* species (Bucklin *et al.*, 1992; Bucklin and Lajeuenesse, 1994; Bucklin *et al.*, 1995; Bucklin *et al.*, 1996; Bucklin, 1998), have been approached via analysis of mitochondrial 16S rRNA gene sequences. Bucklin *et al* (1995) demonstrated that although the species of *Calanus* exhibit exceptional morphological similarity, they are quite distinct in genetic character, and sequences of the

mitochondrial large subunit (16S) ribosomal DNA gene unambiguously discriminate *C. finmarchicus*, *C. glacialis*, *C. marshallae*, *C. helgolandicus*, *C. pacificus*, *C. sinicus*, and *C. hyperboreous*.

It is thought that because of maternal inheritance, mtDNA traits may be better indicators of population structure for marine plankton than nuclear markers. This is because genetically identical mothers and offspring can be more easily recognised even if they are dispersed by periodic mixing (Bucklin *et al.*, 1992). More recent studies (Bucklin *et al*, 1995; Bucklin *et al.*, 1996; Bucklin and Kocher, 1996; Bucklin and Wiebe, 1998) have suggested that the 16S rDNA is too conserved for use as a population genetic marker for *Calanus* species, on all but a large geographical scale. Research has also turned to the sequence analysis of the mitochondrial cytochrome oxidase I (COI) gene. The mitochondrial COI gene sequence has proven useful in enabling the differentiation of *Pseudocalanus* sympatric sibling species, using a molecular protocol based on allele-specific amplification of part of the mitochondrial COI gene by PCR (Bucklin *et al.*, 1998).

1.4.2 Potential molecular applications

Molecular analysis of marine zooplankton is likely to continue to reveal genetic structuring of taxonomic and evolutionary significance. However, one of the perpetual problems in investigating marine zooplankton communities, especially copepod communities, involves the positive identification of the different species present. The use of *C. helgolandicus* for molecular developmental studies demonstrated the need for an accurate and unambiguous technique to distinguish *C. helgolandicus* from other species, especially *C. finmarchicus*. *Calanus* species are reproductively isolated, but the manner in which this is attained results in extreme morphological similarity (Fleminger and Hulseman, 1977; Bucklin *et al.*, 1995). Hence, diagnostic morphological characteristics of *Calanus* species are restricted essentially to minor variations in their secondary sex

characteristics, presenting a persistent problem in the identification of individuals to species level, especially for immature animals. Historically, the geographic location of collection has, at least partially, been relied upon as an indicator of species identity, but the dangers associated with such assumptions are obvious.

Sequencing mitochondrial rDNA has proven that *Calanus* species can be discriminated by their genetic composition, although the process of acquiring the nucleotide sequence is laborious and expensive. In view of the persistent species identification problem experienced by *Calanus* ecologists, there is evidently an essential requirement for a molecular technique with which to identify *Calanus* species routinely. An unambiguous method of species identification is the keystone to any ecological study and such a technique would allow a more accurate determination of *Calanus* temporal and spatial distribution.

1.5 Aims and objectives

The main aim of this study was to develop and apply molecular techniques to increase understanding of the development and systematics of the important copepod genus, *Calanus*. In particular, to ascertain the developmental condition of eggs for *Calanus* egg viability and recruitment measurements, and to identify the morphologically similar *Calanus* species in the North Atlantic. With respect to the molecular techniques applied to the acquisition of *Calanus* developmental genes, the work was performed on *C. helgolandicus*, the *Calanus* species dominant in the coastal waters off Plymouth (UK).

The specific research objectives were as follows:-

Objective I) To construct a *Calanus helgolandicus* genomic library, providing a potential source of any *Calanus* genes (Chapter 3)

Objective II) To obtain *Calanus helgolandicus* developmental genes expressed in early embryogenesis (Chapters 4 and 5)

Objective III) Elucidate the expression pattern of developmental genes to establish their suitability as an egg viability probe (Chapters 4 and 5)

Objective IV) Develop a molecular technique to distinguish between morphologically similar *Calanus* species in the North Atlantic (Chapter 6)

Objective V) Demonstrate the integrity of the developed molecular *Calanus* species identification technique by applying it to a wide range of geographical samples (Chapter 7)

CHAPTER 2

Materials and Methods

CHAPTER 2

Materials and Methods

2.1 Materials

2.2 Microbiological methods

2.2.1 Media for microbiology

2.2.2 Handling and culturing *E. coli*

2.2.3 Preparation of electrocompetent cells

2.2.4 Preparation of chemically competent cells

2.2.5 Transformation of electrocompetent *E. coli* cells

2.2.6 Transformation of chemically competent *E. coli* cells

2.2.7 Preparation of ampicillin stock solution

2.2.8 Lac selection of plasmids

2.2.9 Storage of transformed *E. coli* cells

2.3 Nucleic acid methods

2.3.1 Buffers and solutions for molecular biology

2.3.2 Phenol extraction of DNA

2.3.3 Phenol/chloroform extraction of DNA

2.3.4 Chloroform extraction of DNA

2.3.5 Ethanol precipitation of DNA

2.3.6 Electrophoresis of DNA

2.3.7 Recovery of DNA from agarose gels

2.3.8 Preparation of the vector pBluescript SK^- for molecular cloning

2.3.9 Treatment of DNA to create blunt termini

2.3.10 Ligation of blunt-ended PCR product to digested pBluescript SK⁻

2.3.11 Identification of recombinant clones by colony PCR

2.3.12 Recovery of recombinant plasmid DNA from *E. coli*

2.3.13 Preparation of single stranded DNA for sequencing using Dynabeads

2.3.14 Sequencing of single stranded DNA using the chain termination method

2.3.15 Preparation of acrylamide gel for sequencing

2.3.16 Sequencing gel electrophoresis

2.3.17 Autoradiography

2.3.18 Southern transfer of DNA to Hybond™-N membranes

2.1 Materials

Enzyme	Supplier
DNA ligase	Promega UK Ltd
	Stratagene
Dynazyme	Flowgen Instruments Ltd
Pfu DNA polymerase	Stratagene
Restriction endonucleases	Promega UK Ltd
Sequenase 2.0™	Amersham LIFE SCIENCES
Taq DNA polymerase	Promega UK Ltd

Chemical	Supplier
40% (w/v) Acrylamide/	Anachem Ltd
Bis-acrylamide (19:1)	
Agarose	Promega UK Ltd
Aquaphenol	Appligene Oncor
Bactotryptone	Difco Laboratories
Bromophenol blue	Sigma Chemical Co Ltd
Custom oligonucleotides	British Bio-technology
	Products Ltd
	MWG–Biotech UK Ltd
Dimethyldichlorosilane	Fisons FSA Laboratory supplies
Dynabeads® M-280 Streptavidin	Dynal Ltd
Glycerol	Sigma Chemical Co Ltd
Hybond™-N membranes	Amersham LIFE SCIENCES
IPTG (isopropyl-thiogalactoside)	Sigma Chemical Co Ltd
Metaphor agarose gel	FMC Ltd

Qiaex II gel extraction kit	Qiagen
Quick blot paper	Sigma Chemical Co Ltd
RainX	Halfords, Plymouth
[α-^{35}S] labelled dATP	ICN Pharmaceuticals Inc
TEMED	Gibco BRL
Urea	Promega UK Ltd
Whatman 3MM filter paper	Whatman International Ltd
X- gal (5-bromo-4-chloro-3-indolyl-β-D-Galactopyranoside)	Promega UK Ltd
Xylene cyanol	Sigma Chemical Co Ltd
Yeast extract	Difco Laboratories

All reagents were of molecular biology grade and were obtained from MERCK, unless listed above. All *Escherichia coli* (*E. coli*) strains used were K-12 derivatives. The purity of MilliQ water was 18.2 mΩ.

2.2 Microbiological methods

2.2.1 Media for microbiology

LB broth

Bactotryptone	10 g L^{-1}
Yeast Extract	5 g L^{-1}
NaCl	10 g L^{-1}

Adjusted to pH 7.0 with NaOH or HCl

For LB agar add 15g L^{-1} agar.

SOB media

Bactotryptone	20 g L^{-1}
Yeast extract	5 g L^{-1}
NaCl	0.5 g L^{-1}

When dissolved add 2 mL of 1M KCl

pH to 7.0 with NaOH

SOC media

SOC media = SOB media + 1/200 volume 2 M $MgCl_2$
+ 1/50 volume 1M glucose

2.2.2 Handling and culturing *E. coli*

For plasmid cloning the *Escherichia coli* (*E.coli*) strain XL1-Blue, MRF' was routinely used.

E.coli were handled using standard aseptic techniques as described by Sambrook *et al.*, (1989). Isolation of single *E. coli* colonies was achieved by streaking cells on a LB agar plate with ampicillin selection where appropriate. The plate was then incubated inverted at 37°C overnight. Colonies on LB agar were stored for 2 or 3 months at 4°C with the Petri dish sealed with parafilm. Cultures were routinely grown by inoculating LB broth, in universal tubes for small cultures (5-10mL) and Erlenmeyer flasks for larger cultures (>50mL), with a single isolated colony. The cultures were grown at 37°C with shaking at 225rpm and ampicillin selection where appropriate. Cell densities were monitored spectrophotometrically using OD_{600} values.

2.2.3 Preparation of electrocompetent *E. coli*

A single *E. coli* colony was picked and streaked onto a LB agar plate. The plate was incubated at 37°C overnight. A single colony was inoculated from this plate into 10

mL LB broth and shaken (225 rpm) at 37°C overnight. A 5 mL aliquot of this preculture was used to inoculate 200 mL prewarmed (37°C) LB broth in a 500 mL flask and was shaken at 37°C until A_{600} = 1.0. The resultant log phase *E. coli* culture was treated as follows to make the cells competent.

The cell culture was cooled on ice and 50 mL aliquots transferred to precooled Falcon 50 mL BlueMax tubes. The cells were pelleted in a precooled centrifuge at 4°C for 5 min (RCF = 2500) and re-suspended in 25 mL precooled (4°C) MilliQ water. The wash step was repeated by re-pelleting the cells at 4°C for 5 min (RCF = 2500) and re-suspending in 25 mL MilliQ water. The cells were finally pelleted at 4°C for 5 min (RCF = 2500) and the resulting pellets pooled.

The final volume of pelleted cells was checked and sterile glycerol added to a final concentration of 10%. For storage the cells were snap frozen in 100 µL aliquots in liquid nitrogen and stored at –80°C for later use.

2.2.4 Preparation of chemically competent *E. coli*

A single colony of *E. coli* was streaked onto a LB agar plate and incubated at 37°C overnight. A single colony from this plate was used to inoculate a 5 mL LB broth preculture which was grown at 37°C overnight with shaking (225 rpm). A 1 mL aliquot of this preculture was used to inoculate 50 mL of prewarmed (37°C) LB broth in a 500 mL flask and was shaken (225 rpm) at 37°C until the cells reached exponential phase (OD_{600} = 0.6–0.8). The entire 50 mL culture was used to inoculate 250 mL prewarmed LB broth, and was incubated with shaking (225 rpm) at 37°C until A_{600} = 0.6 – 1.0.

The culture was transferred to four chilled Oakridge centrifuge tubes (Nalgene), and incubated on ice for 15-20 min. The cells were pelleted at 0°C for 4 min (RCF = 2500) and washed with 25 mL per tube of ice cold transformation buffer I (TfBI).

TfB I: 100mM $RbCl_2$
50mM MnCl
35mM KOAc
10 mM $CaCl_2$
15% (v/v) glycerol
Adjusted to pH 5.8 with 1M acetic acid.
Filter sterilised.

The cells were then pooled, divided into two tubes and centrifuged at 0°C for 4 min (RCF = 2500). The pellets were re-suspended in 50 mL TfBI per tube and incubated on ice for 30 min. The cells were then re-pelleted at 0°C for 4 min (RCF = 2500) and the pellet re-suspended in 6 mL ice cold transformation buffer II (TfBII) per tube.

TfBII: 10 mM MOPS
75 mM $CaCl_2$
10 mM $RbCl_2$
15% (v/v) glycerol
Adjusted to pH 6.8 with 1M KOH.
Filter sterilised.

The cells were snap frozen in 0.5 mL aliquots in liquid nitrogen and stored at –80°C.

2.2.5 Transformation of electrocompetent *E. coli* cells

Electrocompetent *E. coli* cells were thawed on ice and aliquots of 40 µL – 80 µL of the cell suspension dispensed into precooled sterile 2 mm electroporation cuvettes (Flowgen) maintained on ice. 1 µL – 5 µL of ligated DNA was added to each cuvette, gently mixed and incubated on ice for 1 min. The cells were then electroporated in a Cellject electroporation system (Flowgen) by applying 12.5 KV/cm of current with a time pulse of 5 ms. The system was set with the following parameters:

Voltage	2500 V
Resistance	132 Ω
Capacitance	40 μF

Immediately following electroporation 1 mL of SOC media was added. Electroporated cells in SOC were transferred into a 15 mL Falcon 2059 polypropylene tube and incubated at 37°C with shaking (225 rpm) for 1 h. The transformed cells were then spread in 250 μL aliquots onto four LB agar plates, containing ampicillin for plasmid selection. The plates were incubated at 37°C overnight or until visible colonies had formed.

2.2.6 Transformation of chemically competent *E. coli* cells

Competent *E. coli* were removed from storage at –80°C and thawed on ice. The thawed cells were gently mixed and 100 μL aliquots placed into precooled 15 mL Falcon 2059 polypropylene tubes. Between 1 μL and 5 μL of ligated DNA was added to each aliquot of cells and mixed gently. The transformation mixture was incubated on ice for 30 min, followed by a 45 sec heat shock at 42°C in a water bath. The mixture was then returned to ice for 2 min and 900 μL SOC (supplemented with a final concentration of 10 mM $MgSO_4$) added. The cells were shaken (225 rpm) at 37°C for 1 h prior to being spread in 250 μL aliquots onto LB agar plates containing ampicillin for plasmid selection. The plates were incubated at 37°C overnight.

2.2.7 Preparation of ampicillin stock solution

Ampicillin was dissolved in MilliQ water to a final concentration of 50 mg mL^{-1} followed by filter sterilisation. The stock solution was stored at –20°C.

For use in LB broth or LB agar the stock solution was added to autoclaved media below 60°C to a final concentration of 100 μg mL^{-1}.

2.2.8 Lac selection of plasmids

Agar plates were thoroughly dried in a laminar flow hood and 20 μL of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) and 2 μL of isopropyl-thiogalactoside (IPTG) per plate were mixed and spread over the surface of each plate. The plates were again dried before transformed *E. Coli* cells were spread over the surfaces. Following overnight incubation of plated cells at 37°C, the X-gal plates were placed at 4°C for up to one hour to assist the blue colour of the non-recombinants to develop.

2.2.9 Storage of transformed *E. coli* cells.

For short term storage, recombinant *E. coli* cells were streaked onto LB agar plates with ampicillin selection. The plates were incubated overnight at 37°C, wrapped in parafilm and stored at 4°C for up to a few weeks.

For long term storage, a single colony was used to inoculate 5 mL of LB broth containing ampicillin, and shaken at 37°C (225 rpm) for 12 h to 16 h. A 0.5 mL aliquot of this culture was added to a sterile cryovial containing 0.5 mL 30% sterile glycerol (final concentration 15% glycerol). The cells were immediately snap frozen in liquid nitrogen and then placed at –80°C for long term storage.

2.3Nucleic acid methods

2.3.1 Buffers and solutions for molecular biology

B + W Buffer	Denaturation solution
2.0 M NaCl	0.5 N NaOH
1 mM EDTA	1.5 M NaCl
10 mMTris-HCl pH 7.5	

Neutralisation solution
(Southern transfer)
0.5 M Tris-HCl, pH 7.5
3 M NaCl

20 x SSC
3 M NaCl
0.3 M Sodium citrate

50 x TAE buffer (pH 7.6)
2 M Tris-HCl pH 7.6
2 M Acetic acid
50 mMEDTA

10 x TBE buffer
89 mM Tris-HCl pH 8.3
89 mM Boric acid
2 mM EDTA

TE buffer
10mM Tris-HCl pH 8.0
1 mM EDTA

All the above buffers and solutions were made with MilliQ water and were autoclaved at 115°C and 15 pounds per square inch for 15 min.

2.3.2 Phenol extraction of DNA

Phenol was prepared for DNA extractions by repeatedly extracting with an equal volume of TE buffer until the pH of the aqueous phase was >7.6. 8-hydroxyquinoline was added to a final concentration of 0.1% as an antioxidant and to enhance the colour of the phenol. The saturated phenol was stored in the dark at 4 °C. Alternatively, equilibrated Aquaphenol was purchased from Appligene, Oncor.

To the DNA sample to be extracted an equal volume of phenol was added and the sample mixed by gentle inversion of the tube. The phenol and aqueous phase were separated by centrifugation for 5 min (RCF = 11600) and the upper aqueous phase transferred to a fresh tube.

2.3.3 Phenol/chloroform extraction of DNA

Equal volumes of equilibrated phenol and chloroform:isoamyl alcohol 24:1 were mixed and stored ready for use at 4°C. An equal volume of phenol/chloroform:isoamyl alcohol 24:1 was added to the DNA. The sample was gently mixed, separated by centrifugation for 5 min (RCF = 11600) and the upper aqueous layer transferred to a clean tube, leaving the interface intact.

2.3.4 Chloroform extraction of DNA

An equal volume of chloroform:isoamyl alcohol 24:1 was added to a suitable volume of DNA solution. The two phases were mixed by inverting the tube or briefly vortexing, and then separated by centrifugation for 5 min (RCF = 11600). The upper aqueous phase was transferred to a clean tube, ensuring no contamination of chloroform.

2.3.5 Ethanol precipitation of DNA

To the DNA solution 0.1 volume of 3M NaOAc and 2.5 volumes of absolute ethanol were added and thoroughly mixed. The solution was incubated either at –20°C overnight, -80°C for >1h, or at –196°C for 5 min. The DNA was pelleted by centrifugation for 10 min (RCF = 11600) and washed with 70% ethanol (precooled to –20°C). The pellet was either air dried or briefly dried in a vacuum desiccator, and re-suspended in an appropriate volume of MilliQ water or TE buffer.

2.3.6 Electrophoresis of DNA

Gel moulds were assembled by sealing the ends with tape and placing the well-forming comb in position. A weighed amount of agarose (Promega UK Ltd) was added to TAE buffer to a final concentration of 0.5 – 1.5 % (w/v) depending on the size of the fragments to be resolved. After heating in a microwave oven until the agarose had

dissolved, the solution was allowed to cool to about 50°C and ethidium bromide solution (10 mg mL^{-1}) added to a final concentration of 0.5 μg mL^{-1}. The solution was poured into the prepared gel mould and allowed to set. When the gel had completely cooled and set the comb was removed and the gel placed horizontally in the electrophoresis apparatus. Sufficient TAE buffer containing ethidium bromide to a final concentration of 0.5 μg mL^{-1} was used to fill the electrode chamber and cover the gel to a depth of 1mm. Electrophoresis was carried out at 5 – 10 V cm^{-1}. Samples were loaded in 1/6 volume 6x loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol, 30% glycerol in MilliQ water). To estimate the size of DNA fragments, samples were co-migrated with appropriate DNA markers. The size of the markers are as follows:-

Bp ladder	100 bp – 1000 bp in 100 bp increments (Promega UK Ltd).
Kb ladder	1 Kb – 10 Kb in 1 Kb increments (Promega UK Ltd).
λ *Hin*dIII	23.1 Kb, 9.4 Kb, 6.6 Kb, 4.4 Kb, 2.3 Kb, 2.0 Kb, 0.5 Kb.
λ *Sma*I	19.4 Kb, 12.2 Kb, 8.6 Kb, 8.3 Kb.
PhiX174/*Hae*III	1.35 Kb, 1.08 Kb, 0.87 Kb, 0.60 Kb, 0.31 Kb, 0.28 Kb, 0.27 Kb, 0.23 Kb, 0.19 Kb, 0.12 Kb, 0.07 Kb.

DNA in the gel was viewed after electrophoresis by ethidium bromide fluorescence on an ultra-violet transilluminator (U. V. P. Inc) and photographed using a Polaroid Camera (GRI) fitted with a UV filter and Polaroid Type 667 positive film.

2.3.7 Recovery of DNA from agarose gels

DNA was recovered from agarose gels using a Qiaex II Extraction kit (Qiagen) according to the manufacturer's instruction. The final elution was performed in 20 μL MilliQ water or TE buffer.

2.3.8 Preparation of the vector pBluescript SK⁻ for molecular cloning

The plasmid pBluescript SK⁻ was prepared for ligation by digesting with a suitable restriction enzyme for which there is only a single restriction site in the polylinker of the plasmid. The restriction enzymes used to linearise the plasmid resulting in blunt termini were *Eco*RV and *Sma*I.

Digest reaction 10μL	
pBluescript SK⁻ (360 ng μL^{-1})	3μL
10x reaction buffer	1μL
*Eco*RV / *Sma*I (8-10 U μL^{-1})	1μL
MilliQ water	5μL

The above reactants were combined in a microcentrifuge tube and incubated for >1 h at 37°C for digests with *Eco*RV, and at 25°C for digests with *Sma*I. If necessary the reaction was stopped by heat inactivating the enzymes at 65°C for 15 min. The plasmid was examined for complete digestion by running on an agarose gel and observed on a transilluminator.

2.3.9 Treatment of DNA to create blunt termini

Purified PCR products (Section 2.3.7) were treated with Pfu DNA polymerase to create blunt ends for ligation with pBluescript SK⁻ as follows:

Purified PCR product	10μL
2 mM dNTP's	1 μL
10x cloned reaction buffer	1 μL
Pfu DNA polymerase ($3U \mu L^{-1}$)	1 μL

The above reactants were combined in a microcentrifuge tube and incubated at 72°C for 30 min. An aliquot of polished PCR product was either used directly for ligation or stored at 4°C until use.

2.3.10 Ligation of blunt-ended PCR product to digested pBluescript SK⁻

PCR fragments prior to cloning were tested to see if they contained a digestion site for either *Eco*RV or *Sma*I. In the absence of a digestion site the fragment was ligated into the corresponding digested plasmid, ie., if no *Eco*RV site was found in the PCR product the fragment would be cloned into an *Eco*RV digested plasmid. This enabled *Eco*RV to be added to the ligation reaction to prevent self-ligation of the plasmid. If both restriction sites were found in the DNA fragment the restriction enzyme was omitted from the ligation reaction and the volume made up with MilliQ water. Ligations were carried out in the approximate molar end ratio of 10:1 insert:vector as follows:

Digested pBluescript SK⁻	1μL (~20 ng)
(1:5 dilution of digest reaction, section 2.3.8)	
10x ligase reaction buffer	1 μL
10 mM rATP	0.5 μL
Blunt-ended PCR product	5.5 μL
*Eco*RV/ *Sma*I (if appropriate)	1 μL
T_4 DNA ligase (3-4U μL^{-1})	1 μL

The ligation reaction was incubated for 2-3 h at room temperature or 15°C overnight. It was used either immediately for transformation with chemically competent cells, ethanol precipitated and re-suspended in MilliQ water for transformation with electrocompetent cells, or stored at 4°C.

2.3.11 Identification of recombinant clones by colony PCR

Recombinant clones were primarily identified by the blue/white system of Lac selection on X-gal LB agar plates (Section 2.2.8). Further identification of recombinant clones was achieved by amplification of the inserted DNA, using primers complementary to regions approximately 200 bp outside the pBluescript SK⁻ polylinker.

The primer sequences were as follows:

BS1F: 5'-AAAGGGGGATGTGCTGCAAGGCG-3'

BS1R: 5'-GCTTCCGGCTCGTATGTTGTGTG-3'

Individual colonies were picked from the transformation plates using a sterile loop. Each colony was re-suspended in 10 μL MilliQ water in a 0.7 mL tube and 1 μL was patched out on a LB agar plate containing ampicillin. The patched colonies were grown at 37°C overnight and stored at 4°C wrapped in parafilm. The re-suspended colonies were boiled for 4 min in a water bath and then placed on ice immediately.

The PCR mix for each colony suspension was made as follows:

2 mM dNTP's	2 μL
10x *Taq* DNA polymerase buffer	2 μL
Primer BS1F (100 ng μL^{-1})	0.5 μL
Primer BS1R (100 ng μL^{-1})	0.5 μL
Taq DNA polymerase	0.2 μL
MilliQ water	4.8 μL

10 μL of the mix were added to the bacterial suspension, mixed well and amplified in a thermal cycler (PTC-100™, MJ Research, Inc.) The cycling parameters included an initial denaturation step of 94°C (5 min) followed by 25 cycles of 68°C (1 min), 72°C (1 min), and 94°C (1 min). A final annealing phase at 68°C (2 min) was followed by an extension phase at 72°C (2 min) and storage at 4°C until use. 5μL aliquots of the ampification reactions were analysed by agarose gel electrophoresis with a control of non-ligated amplified pBluescript SK^- and 100 ng λ *Hin*dIII molecular markers. Plasmids containing inserts were identified by the increased size of the PCR amplification product compared with that of control plasmid.

2.3.12 Recovery of recombinant plasmid DNA from *E. coli*

Promega Wizard miniprep DNA purification system

E.coli were patched onto ampicillin LB agar plates and grown overnight at 37°C. A single colony from the plate was used to inoculate 5 mL of LB broth containing ampicillin, in a sterile universal. The culture was then grown shaking (225 rpm) at 37°C for 12 – 16 h.

The cells were pelleted by centrifugation of 1.5 mL of the culture in a microcentrifuge tube for 2 min (RCF = 11600). The supernatant was removed and replaced with a further 1.5 mL of the same culture and again pelleted by centrifugation for 2 min (RCF = 11600). The pellet was re-suspended in 200 µL of "cell resuspension solution", followed by the addition of 200 µL of "cell lysis solution" and 200 µL of "neutralising solution". The cell suspension was mixed by inverting the tube four times between each addition. Finally the suspension was centrifuged for 5 min (RCF = 11600). For each sample a Wizard minicolumn was prepared by attaching a barrel into a minicolumn and connecting to a vacuum manifold. 1 mL of Wizard miniprep DNA purification resin was added to each barrel. The cell lysate was carefully removed from the pellet and transferred to the prepared column. By applying a vacuum the lysate was pulled through the column and washed with 2 mL "column wash solution". The minicolumn was dried by transferring to a microcentrifuge tube and centrifuging for 2 min (RCF = 11600). The DNA was eluted into a clean collection tube with 50 µL MilliQ water by centrifugation for 20 sec (RCF = 11600).

Following examination of the recovered plasmid DNA by gel electrophoresis the preparations were stored at 4°C short term or –20°C long term.

2.3.13 Preparation of single stranded DNA for sequencing using Dynabeads

Following identification of recombinant clones by colony PCR, the clones were reamplified in preparation for sequencing. For such a reaction one of the primers must be

biotinylated, such that one strand of the PCR product possesses a 5'-biotin label. Each recombinant clone to be sequenced was picked with a sterile loop and re-suspended in 10 μL MilliQ water in a 0.7 mL tube. The colonies were boiled in a water bath for 4 min and then stored on ice. 40 μL of PCR mix, as follows, was added to each bacterial suspension.

2 mM dNTP's	5 μL
10x *Taq* DNA polymerase buffer	5 μL
*Primer BS1F (100 ng μL^{-1})	1 μL
*Primer BS1R (100 ng μL^{-1})	1 μL
Taq DNA polymerase	0.5 μL
MilliQ water	27.5 μL

*One of the primers must be biotinylated

Amplification was performed in a thermal cycler (PTC-100™, MJ Research, Inc.) following the same cycling parameters as for colony PCR (Section 2.3.11). 5 μL aliquots of the ampification reactions were analysed by agarose gel electrophoresis to check amplification efficiency.

DNA was prepared for sequencing using M-280 streptavidin Dynabeads^R (Dynal Ltd). For each DNA template 20 μL of Dynabeads (10 μg μL^{-1}) were prepared by washing. The Dynabeads were aliquoted into a microcentrifuge tube and pulled to one side with a magnetic separation stand (Promega UK Ltd). Dynabead storage buffer was removed and 20 μL of B + W buffer added. The beads were gently re-suspended before being collected with a magnetic separation stand, and the supernatant removed. The beads were then re-suspended in 40 μL of B + W buffer, giving a final bead concentration of 5 μg μL^{-1}.

The washed Dynabeads were used to immobilise the biotinylated PCR products by transferring 40 μL of the 50 μL PCR reaction to a clean tube and adding 40 μL of

prewashed Dynabeads. The 5'-biotin labelled strand of the PCR product was allowed to bind to the streptavidin-coated Dynabeads by incubating for 15 min at room temperature with occasional gentle mixing.

The Dynabead/PCR product complex was washed once with 40 µL B + W buffer and stored in 40 µL B + W buffer at 4°C until ready for use. The DNA duplex was denatured by removal of the supernatant, re-suspension of the beads in 100 µL 0.1 M NaOH and incubation at room temperature for 10 min. Following incubation the non-biotinylated strand of DNA was removed by collection of the Dynabeads and removal of the supernatant. The beads and biotinylated strand were then washed once with 100 µL 0.1 M NaOH, once with 100 µL B + W buffer and twice with 100 µL TE. The supernatant was removed, the beads re-suspended in 7 µL MilliQ water and the single stranded DNA used immediately for sequencing.

2.3.14 Sequencing of single stranded DNA using the chain termination method

Sequencing was achieved following the chain termination method (Sanger *et al.*, 1977) using Sequenase 2.0 sequencing kit (Amersham) as follows. The single stranded DNA separated with streptavidin Dynabeads was annealed to the appropriate sequencing primer. The sequencing primer T3 was used with the biotinylated BS1F strand, and the sequencing primer T7 was used with the biotinylated BS1R strand.

Annealing mixture (10 µL)	
Single stranded DNA	7 µL
5x sequenase reaction buffer	2 µL
Sequencing primer (T3 or T7), (0.5 pmol)	1 µL

The primer was annealed by heating for 2 min at 65°C in a water bath and then allowed to cool slowly to below 35°C over a 15 – 30 min period. The annealed mixture was chilled on ice until use.

Colour coded microcentrifuge tubes were filled with 2.5 µL of each termination reaction ddATP (red), ddCTP (blue), ddGTP (green), ddTTP (yellow). Each termination reaction contained 80 µM deoxynucleotides, 50 mM NaCl and 8.0 µM specific dideoxynucleotide. The 5x labelling mix containing 7.5 µM of each deoxynucleotide was diluted 1 in 5 to a 1x working concentration and the Sequenase polymerase (13 U μL^{-1}) was diluted 8 fold with Sequenase dilution buffer. To the ice cold annealed DNA mixture the following were added:

0.1 M DTT	1µL
Diluted labelling mix	2 µL
[α-^{35}S] dATP	0.5 µL
diluted Sequenase	2 µL

The reactants were mixed and incubated at room temperature for 2-5 min. Following the incubation 3.5 µL of each labelling reaction were transferred to each termination tube (ddGTP, ddATP, ddTTP and ddCTP). The reactions were mixed, left to incubate at 37°C for a further 5 min and then stopped with the addition of 4 µL of stop solution (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol).

2.3.15 Preparation of acrylamide gel for sequencing

A 6% acrylamide / urea solution was made by dissolving 120 g of ultrapure urea in 37.5 mL 40% (w/v) acrylamide/bis-acrylamide (19:1) solution and 25 mL 10 x TBE in a total volume of 250 mL made up with MilliQ water. Once dissolved the solution was passed twice through a Whatman 3MM filter paper and then stored at 4°C for several weeks.

Two glass sequencing plates were washed in hot water with detergent, rinsed and the inside of each plate cleaned thoroughly, first with acetone and secondly with IMS. The inside of the smaller of the plates was cleaned with either rainX or dimethyldichlorosilane solution to aid separation of the plate from the gel. The plates were placed face to face with the clean side innermost and were separated by a 0.35mm spacer at each vertical edge. The plates were taped tightly together along both sides and the bottom and clipped with bulldog clips.

The gel solution was prepared by the addition of 160 µL of 10% ammonium persulphate (APS) and 65 µL TEMED to 45 mL 6% acrylamide solution. The gel was then immediately poured using the barrel of a 50 mL syringe fitted with a 200 µL tip. The well-forming comb was put in place inverted, and the gel left to polymerise for between 3 h and overnight.

2.3.16 Sequencing gel electrophoresis

The clips and tape were removed from the polymerised gel and the gel plates assembled vertically in the sequencing apparatus (Poker Face II, SE 1600, Hoefer Scientific Instruments). The buffer tanks were filled with 1 x TBE, the well-forming comb removed and the loading space washed with 1x TBE in a 20 mL syringe to remove excess urea. The gel was pre-run for between 15 min and 1 h at a limiting power of 50 watts and voltage less than 2000 V.

Prior to loading, the samples were heated to 72°C for 2 min and the sharks-tooth comb was placed into the loading well with the teeth just piercing the top of the gel. Approximately 3 µL of each sample were loaded on to the gel and electrophoresis performed under the same parameters as the pre-run. The time of electrophoresis was determined by the migration of xylene cyanol and bromophenol blue in the stop buffer and the length of sequencing read required.

Post-run manipulation of the sequencing gel involved dismantling the sequencing apparatus and removing the smaller of the plates, leaving the gel attached to the larger plate. The gel was left on the plate in sequencing soak (5% acetic acid, 15% methanol) for 15 – 30 min. The gel was removed from the soak solution and transferred to Whatman 3MM filter paper. The exposed side of the gel was then covered with Clingfilm and the Whatman 3MM filter paper and Clingfilm trimmed closely to the size of the gel. The gel was dried in a heated vacuum dryer (slab gel dryer, SE 1160, Hoefer Scientific Instruments) for between 1 h and 3 h. When dry the Clingfilm was removed from the gel.

2.3.17 Autoradiography

The dried sequencing gel was exposed to X-ray film (Genetic Research Instrumentation) in a sealed cassette overnight at room temperature. The film was subsequently processed under a red filtered light by soaking for 5 min in GBX developer (Sigma) and then transferred to GBX fixative for 5 min (Sigma), with a brief wash in water in between. The film was washed under running water for 10 min prior to air-drying. The exposure time was increased if necessary, according to the intensity of the signal.

2.3.18 Southern transfer of DNA to Hybond™–N membranes

DNA was transferred to Hybond™-N membranes using a modification of the method described by Southern (1975).

The DNA to be transferred was electrophoresed in an agarose gel, and photographed aligned with a ruler (Section 2.3.6). The gel was then rinsed briefly in distilled water and incubated for 15 min in 5 – 10 volumes of denaturing solution, with gentle shaking at room temperature. The incubation was repeated in fresh denaturing solution for a further 15 min. The gel was again briefly rinsed in distilled water and then

incubated twice in neutralising solution with gentle agitation for 15 min at room temperature.

A sheet of Hybond™-N membrane, cut to the exact size of the gel, was pre-wet by floating on 2x SSC until saturated. A container was filled with 20x SSC and a suitable sized gel tray was used inverted as a platform, on which 3 sheets of Whatman 3MM filter paper were placed such that the ends formed wicks into the 20x SSC blotting solution. The gel was placed wells down on the wick and the pre-wet Hybond™-N membrane carefully placed on top with the exclusion of air bubbles. On top of the membrane were layered two sheets of Whatman 3MM filter paper, approximately 15 sheets of quick blot paper (all cut to the size of the gel) and a glass plate and weight. The transfer was allowed to proceed for 2 h – overnight.

After transfer the apparatus was carefully dismantled and the position of the wells marked in pencil on the membrane, along with its orientation with the gel. The membrane was washed in 6x SSC for 5 min at room temperature, briefly air-dried, baked at 80°C for 2 h and exposed to 312 nm UV light for 1.5 min on a transilluminator. The membranes were stored either dry under vacuum at room temperature, or sealed with 2x SSC in a bag at 4°C until ready for prehybridisation.

CHAPTER 3

Construction of a Genomic Library from *Calanus helgolandicus*

CHAPTER 3

Construction of a Genomic Library from *Calanus helgolandicus*

3.1 Introduction

3.1.1 Strategies for gene isolation

3.1.2 What type of library?

3.1.3 Which type of vector?

3.2 Methods

3.2.1 Overview of genomic library construction

3.2.1.1 Preparation and manipulation of DNA for genomic library construction

3.2.1.2 Lambda GEM-12

3.2.1.3 Lambda DASH II

3.2.1.4 Amplifying the library

3.2.2 DNA extraction from *Calanus helgolandicus*

3.2.2.1 Basic extraction procedure

3.2.2.2 Purification of extracted DNA by density gradient centrifugation

3.2.3 Partial restriction endonuclease digestion of genomic DNA

3.2.4 Assembly and titration of intact recombinant λ DASH II bacteriophage

3.2.4.1 Ligation of partially digested genomic DNA into λ DASH II

3.2.4.2 Preparation of host bacterial strain

3.2.4.3 Packaging and titration of recombinant λ DASH II

3.2.5 Dephosphorylation and size fractionation of partially digested DNA

3.1. Introduction

The vital role played by developmental genes in the pathway of a fertilised egg developing into an adult make them of considerable scientific interest. In addition to the general scientific interest in developmental genes, their identification and analysis will enable a better understanding of the reproductive and developmental biology of *Calanus* species, which often dominate zooplankton biomass (Jaschnov, 1970; Raymont, 1983). By beginning to understand the mechanisms of *Calanus* development we can potentially investigate the effect of environmental factors on the expression of key developmental genes, and utilise molecular information obtained, to address the problematic measurement of copepod production.

3.1.1 Strategies for gene isolation

With the development within the last two decades of novel molecular techniques in the biomedical field, there are numerous methods available for the isolation of new genes. However, the application of these molecular techniques to novel systems, including marine systems, is not trivial.

A simple and routine technique for acquiring gene sequences is the use of the polymerase chain reaction (PCR) (Saiki *et al.*, 1985; 1988). From minute quantities of starting material any specific sequence of DNA can be amplified provided that the sequences of the flanking regions are known. It is the lack of sequence data in biological oceanography that inhibits the design of specific oligonucleotides and hence, limits the application of PCR. As discussed by Falkowski and LaRoche (1991) very few genes from marine organisms have been studied in detail, but it is often assumed that conserved amino acid regions in well-characterised proteins are conserved in homologous proteins of marine organisms. This is certainly not always the case, illustrated clearly by the rubisco enzyme, which although ubiquitous among autotrophs is not highly conserved between the different

algae groups (Keen *et al.*, 1988). This lack of conservation at the amino acid level exemplifies the problem of seeking conservation at the DNA level for oligonucleotide design.

In many cases the lack of sequence data for marine organisms, from which PCR primers can be designed, may be overcome by the use of oligonucleotides of degenerate sequence. With ready access to gene sequences on the computer database [European Molecular Biology Laboratory (EMBL) database; using the Genetics Computer Group Sequence Analysis Software Package, Devereux *et al.*, 1984] it is possible to align homologous genes from many different organisms. If suitably conserved regions can be identified oligonucleotide primers may be designed with degeneracies to take into account variation between certain base pairs of the homologous sequences. However, the design of degenerate oligonucleotides for marine systems is impeded by the lack of any codon usage tables for marine organisms.

For the study of *C. helgolandicus* developmental genes the design of oligonucleotide primers from aligned homologous developmental genes has not proved possible because there is neither enough sequence information from different organisms nor enough homology between genes for the design of suitable PCR primers. The lack of homology between these developmental genes may be explained by the fact that they code for transcription factors rather than for enzymes. Hence, selection pressure on such genes may be restricted to a small region encoding the DNA binding site.

In the absence of two disparate sites within a gene from which to design primers for amplification, it is sometimes possible to amplify a smaller fragment from the gene, using primer sites within a single conserved domain. This strategy has indeed been used in *C. helgolandicus*. Degenerate primers have been derived from sequence comparisons of a highly conserved amino acid sequence motif, known as the homeodomain, in homologous developmental genes cloned from various species (Smerdon, 1998). These primers have

been successfully utilised on *C. helgolandicus* genomic DNA to identify specific regions of developmental genes known as homeoboxes (Smerdon, 1998). The homeobox is a region of DNA, 180 bp in length, common to many developmental genes and conserved throughout evolution. Although it is possible to use sequence comparisons within existing databases in order to obtain an idea of the homeobox's identity, for an accurate prediction it is necessary to acquire sequence information of the flanking regions. This problem is not specific to homeoboxes, but is universal to the strategy of amplifying a small region of DNA within a single conserved domain. Hence, the homeobox sequence alone is not sufficient to constitute the isolation of a gene, although it does provide a powerful tool with which to probe a prepared clone library (Chapter 4).

In the event that the rapid, sensitive and relatively simple technique of PCR should fail to provide a suitable procedure with which to obtain *Calanus* developmental genes, more complex cloning procedures must be adopted. Since such genetic engineering procedures have been in existence for over two decades it may be thought that the basic steps are now commonplace. To a certain degree this is true. However, the complex procedure of precisely manipulating living material must not be underestimated. This realisation is summarised by Gannon and Powell (1995) "It is frequently a surprise to those entering the field (of molecular biology) that the experimental procedures involved, which have been analysed to an extent that they no longer warrant mention, are far from trouble-free. It should be reiterated that the skills and care of a biochemist are required to purify DNA free from all biological and organic factors which can interfere with subsequent steps, that the enzymes used are delicate tools which demand correct handling and that the biological hosts which serve as recipients for the foreign DNA require the knowledge and experience of a trained microbiologist." However, with these considerations taken into account molecular cloning can provide an excellent basis for gene analysis.

3.1.2 What type of library?

In the main there are three ways by which a desired gene may be obtained by molecular cloning. 1:- The cloning of a single or small number of DNA fragments understood to contain the gene of interest directly from the purified nucleic acid content of the host tissue. 2:- The construction of a cDNA (complementary DNA) library, derived from the mRNA present at a particular stage or in a particular tissue. 3:- The construction of a genomic library, derived from the genomic DNA of the organism under study, and consisting of a collection of clones sufficient in number to represent the whole genome of that organism.

If a single restriction fragment known to contain the gene of interest, through Southern transfer and subsequent probing, could be separated by electrophoresis it would be a relatively simple task to clone it into a plasmid. Unfortunately this is rarely the case in the initial stages of gene isolation. Certainly it is of little use in isolating a gene from genomic DNA as the genome is of such a size that restriction fragments appear as a long smear and cannot be separated by electrophoresis. It is feasible to clone a portion of this smear containing a selection of DNA fragments including the gene of interest. However, the amount of restricted genomic DNA that can be run, and the sensitivity of a heterologous probe in detecting a single copy number gene in a Southern transfer, means that the ability to locate a specific gene by this method is often impossible. The construction of such a partial library is therefore limited to situations whereby multiple copies of a gene are present in a given genome. This is not normally the case for genes involved in developmental regulation.

A cDNA library is representative of the mRNA present in a given tissue at a given time, and can prove very useful for obtaining sequence information on the coding regions of genes. However, when cloning developmental gene sequences from *Calanus* there is the disadvantage that there is very little starting material available from which to isolate

mRNA. Perhaps even more problematic is that developmental genes are subject to temporal expression, necessitating the construction of libraries from several different stages to ensure that a particular gene transcript is represented within one of these libraries. An additional concern is that the instability of the mRNA during preparation and conversion to cDNA will lead to incomplete clones that represent only parts of the transcripts from which they are derived (Brown, 1995).

The third way of isolating particular genes is through the construction of a genomic library. Such a library must necessarily contain an extensive number of clones in order to represent the whole genome. The exact number depends on the size of the genome of the organism, the required probability of a particular gene being present, and the average size of DNA cloned. This is discussed further in Section 3.2. The disadvantages of a genomic library include the complexity of its construction and the difficult task of searching through the numerous clones for the gene of interest, a large proportion of the clones being non-coding. However, a genomic library does provide an unequivocal starting point, containing all genes in the entire genome of *C. helgolandicus*. This provides an excellent facility with which to study the complex interactions of developmental genes or in fact, any *C. helgolandicus* molecular genetic system. For these considerations it was decided that a genomic library should be constructed as a step towards the analysis of *C. helgolandicus* developmental genes.

3.1.3 Which type of vector?

As mentioned above, a genomic library must consist of a large number of clones in order to represent the whole genome. An important factor reducing the number of clones needed is the size of insert that can be cloned into the chosen vector. For this reason plasmid vectors are seldom used for the construction of genomic libraries, predominantly because obtaining consistently large inserts, when transformation favours smaller

plasmids, is impossible. Routinely, vectors used for genomic libraries are either lambda (λ) bacteriophages or cosmids, because of their capacity for accommodating large inserts. Cosmids consist of the λ *cos* site inserted into a plasmid. They have the advantage that they are capable of accommodating larger inserts than λ, up to 45 kb (Collins and Hohn, 1978), but the disadvantage is that they are more complex both to construct and screen. Perhaps the biggest advantage of λ phage over cosmids as vectors for library construction is that the former produces a higher yield of recombinants per microgram of starting DNA, and is therefore the preferred choice when the amount of starting material is limited.

Having decided to adopt λ bacteriophage as the vector of choice there still remains the selection of which λ bacteriophage to use. "Insertion" vectors have a single target site for insertion of exogenous DNA. However, they have proven more useful for the cloning of small DNA fragments (Pouwels, 1995), since the overall size of the phage DNA which can be packaged is limited to approximately 50 kb. 'Replacement' vectors on the other hand have a pair of sites flanking a segment of non-essential bacteriophage λ DNA, allowing the 'replacement' of this DNA with foreign DNA. For lytic growth of the bacteriophage only about 60% of the viral genome is necessary, and the central third (known as the stuffer fragment) is superficial. Removing the stuffer fragment results in a separate left and right arm between which the foreign DNA can be cloned. In order to be packaged into the λ heads DNA must contain two *cos* sequences in the correct orientation at a distance that corresponds to 78% to 105% of the wild-type λ genome (Sambrook *et al.*, 1989), permitting insert sizes ranging from 9 kb to 23 kb. Thus packaging of recombinant DNA molecules into phage heads provides an excellent size selection system for recombinant bacteriophages.

The reasons mentioned above have led to the widespread use of λ bacteriophage vectors for the construction of libraries in *E. coli*. Two such λ replacement vectors (λ

GEM-12 and λ DASH II) were selected for the *C. helgolandicus* library construction and are discussed below. The precise methodology is given only for the vector that proved most suitable for construction of a *C. helgolandicus* genomic library (λ DASH II).

3.2 Methods

3.2.1 Overview of genomic library construction

3.2.1.1 Preparation and manipulation of DNA for genomic library construction

The first requirement for the construction of a genomic library is the extraction of clean, high molecular weight genomic DNA (Section 3.2.2). Once high molecular weight DNA has been successfully isolated, it is necessary to reduce the fragment size range by restriction enzyme digestion. To help ensure that sequences are represented with equal probability the genomic library should be prepared by partially restricting the DNA with a suitable restriction endonuclease such as *Sau*3AI. Partial digestion with a frequent cutter has the advantage of cutting the DNA as randomly as possible to a suitable size to be cloned. Restriction digest conditions are optimised on a small scale to determine the correct enzyme concentration to generate the desired size range of fragments. The assayed restriction digests are separated by agarose gel electrophoresis and the gel is observed by UV transillumination to determine the amount of enzyme required to produce the maximum intensity of fluorescence in the desired size range. The intensity of fluorescence is related to the mass distribution of the DNA. To obtain the maximum number of molecules in the desired size range for library construction, half the amount of enzyme that produces the maximum amount of fluorescence is used (Seed *et al.*, 1982). These conditions are then used on a large scale to produce the necessary amount of restricted genomic DNA.

To increase the 'efficiency' of the library titre the sample DNA can be further treated in such a way as to prevent multiple inserts. A wide size range of DNA fragments can

result from the *Sau*3AI partial digest and, in the case of the vector system finally used (λ DASH II), there is no method of preventing cloning of randomly co-ligated pieces of DNA. This is problematic because firstly it creates clones containing ligated DNA fragments that are not contiguous in the genome, and secondly these ligated and cloned DNA fragments may exceed the capacity of the packaging system. Similarly, the presence of small DNA fragments in the partial digest will allow ligations resulting in phage genomes too small for packaging. Although these recombinants will not be packaged they will lower the number of vector arms available. The co-ligation of DNA pieces can be prevented by jointly dephosphorylating the partially restricted DNA with calf intestinal alkaline phosphatase (CIAP), and by size fractionation on a sucrose gradient (Section 3.2.5).

3.2.1.2 Lambda GEM-12

The first λ replacement vector considered for the construction of a genomic library was the multifunctional cloning vector lambdaGEM-12, a derivative of EMBL3. It is commercially available from Promega UK Ltd, and is designed for simplified genomic library construction with an ultra-low background of non-recombinants.

The lambdaGEM-12 vector is prepared by removing the stuffer fragment with *Xho*I and then partially filling in the *Xho*I restriction sites. The partially filled-in, dephosphorylated *Xho*I half-site arms are then ligated to partially filled-in *Sau*3AI-digested genomic DNA. The primary ligation products are theoretically single copies of genomic inserts with appropriate arms, since the partial fill-in should prevent self-ligation reactions of vector arms and genomic fragments. This cloning strategy also obviates the need for size-fractionation or alkaline phosphatase treatment of the restricted genomic DNA. LambdaGEM-12 *Xho*I half-site arms were the original choice of cloning vector because of the above mentioned qualities and the requirement for small amounts of starting material.

Successfully extracted high molecular weight DNA from *C. helgolandicus* (Section 3.2.2.1) was partially restricted with *Sau*3AI. The first two nucleotides of the *Sau*3AI site were then filled-in using Klenow (the polymerase portion of DNA polymerase I), and the filled-in genomic DNA was ligated to the lambdaGEM-12 *Xho*I half site arms. To optimise the ligation conditions the molar ratio of lambda DNA to genomic DNA varied between approximately 1:2.5 and 1:0.5. The recombinant λ DNA was then packaged *in vitro* using a phage-infected *E. coli* cell extract to supply the mixture of proteins and precursors required for encapsidating λ DNA. The packaged phage was transfected into the *E. coli* host strain LE-392, mixed with TB top agar and spread onto LB agar bottom plates.

Positive controls (supplied by the manufacturer), including a positive ligation control of 16 kb plasmid insert (pTI11) predigested with *Bam*HI and partially filled-in to produce ends compatible with the half-site arms, and a positive transformation control (λc1857 *Sam*7 DNA), successfully produced high numbers of plaques. However, cloning of the sample DNA repeatedly proved unsuccessful.

The quality of DNA used was the first step to be addressed for potential improvement and a number of methods to purify further the genomic DNA starting material were attempted. *C. helgolandicus* genomic DNA was extracted using a Nucleon Bacc 1 Kit (Scotlab) instead of the normal extraction method (Section 3.2.2). The DNA was also cleaned, following a normal extraction, both with the use of a Qiagen-tip 5 column (Qiagen), and a modified Wizard mini-prep system (Promega UK Ltd.). In preference to the standard phenol extraction following restriction with *Sau*3AI an alternative method of heat inactivating the *Sau*3AI prior to ethanol precipitation was used to eliminate the possibility that phenol contamination was hindering the fill-in reaction. DNA was also purified through a NAP™-5 column (Pharmacia LKB Biotechnology) following digestion, to ensure that the restricted DNA was completely desalted. Unfortunately none of these alterations to the protocol would yield plaques of recombinant

phages. The addition of a small quantity of ligated *C. helgolandicus* DNA to both the positive ligation control and the positive transformation control led to no reduction in titre, suggesting that in fact the problem did not lie with the quality of genomic DNA.

Further factors were tested to try and determine the cause of the problem. These included the use of new and differing concentrations of T_4 DNA ligase, the use of fresh nucleotide buffer for the fill-in reaction, containing fresh dATP, dGTP and acetylated BSA and the supplementing of the ligation reaction with rATP. The ligation step was further investigated by purifying filled-in *C. helgolandicus* DNA through a Wizard clean up column prior to ligation.

The problem was finally isolated to the partial fill-in of the digested genomic DNA. Test ligations including the self-ligation of digested genomic DNA, digested plasmid (pBluescript SK^-) DNA, digested/filled-in genomic DNA, digested/filled-in plasmid DNA and a negative control for each (absence of T_4 DNA ligase), indicated that the partial fill-in was unsuccessful and unable to prevent self-ligation as expected. The use of LambdaGEM-12 *Xho* I half site arms was discontinued and a new approach and vector chosen.

3.2.1.3 Lambda DASH II

A second vector was chosen without the requirement of partially filling in the digested genomic DNA. The chosen vector, lambda DASH II (Stratagene), is also a replacement vector used for cloning large fragments of genomic DNA. The lambda DASH II system takes advantage of spi (sensitive to P2 inhibition) selection. λ phage containing active *red* and *gam* genes (spi^+) cannot normally infect *E. coli* cells that already possess an integrated form of a related prophage called P2. The *red* and *gam* genes in the lambda DASH II DNA are located on the stuffer fragment, therefore removal of the stuffer fragment and insertion of new DNA causes a change from spi^+ to spi^-. Thus, only

recombinants can infect *E. coli* strains lysogenic for P2, whereas the wild-type lambda DASH II phage cannot grow on host strains that contain P2 phage lysogens.

Unlike the lambda GEM-12 half-site arms the new λ DASH II vector has no direct properties for preventing self-ligation of the insert DNA or of the vector arms, apart from the size selection enforced by packaging. The vector however, was double digested with *Bam*HI and *Xho*I to prevent religation of the stuffer fragment to arms.

3.2.1.4 Amplifying the library

Once constructed it is desirable to amplify the library to make a large, stable quantity of high-titre stock. However, it is not recommended to perform more than one round of amplification, since slower growing clones may be significantly under represented, and an ideal genomic library should be random, with all sequences being represented with equal probability. To enable a successful amplification of the library the potential for non-representation of a given sequence can be minimised by using the following equation (Clarke and Carbon, 1976).

$$N = \frac{\ln (1 - P)}{\ln (1 - \frac{1}{n})}$$

where: N = the number of independent recombinants

P = probability of the presence of any given sequence in the genomic library

n = size of the genome relative to a single cloned fragment

This equation calculates the number of clones needed to ensure the representation of any given sequence at a chosen probability. Therefore the library can be amplified to provide a large quantity of high-titre stock which can be stored indefinitely.

3.2.2 DNA extraction from *Calanus helgolandicus*

3.2.2.1 Basic extraction procedure

For the construction of a genomic library, clean high molecular weight genomic DNA is required.

Calanus helgolandicus were collected by net tows (WP2 net, mesh 200 µm, UNESCO 1968) from a coastal station approximately 10 km off Plymouth (western English Channel: 50°15' N; 4°13'W) and transported back to the laboratory within 2 h of collection. Individuals were identified by microscopy and placed in a 50 mL beaker containing filtered sea water. Prior to DNA extraction the animals were concentrated by washing into a plexiglass tube with 200 µm mesh bottom, rinsed briefly with distilled water and aliquots of 50 animals placed in microcentrifuge tubes.

To each microcentrifuge tube containing 50 individuals, 340 µL of homogenising solution (400mM Tris-HCl pH 8, 60 mM EDTA, 150 mM NaCl, 1% SDS) and 10 µL (10 mg mL^{-1}) RNase were added. The copepods were homogenised using a pellet pestle homogeniser (Anachem Ltd.) and incubated at 37°C for 30 min. Following the incubation, proteinase K was added to a final concentration of 250 µg mL^{-1} and the homogenate further incubated at 37°C for 30 min. 100 µL of 5 M Na-perchlorate was added to each tube and incubated at room temperature for 20 min with shaking followed by 20 min at 65°C. The homogenate was then extracted once with phenol/chloroform:isoamyl alcohol 24:1 and once with chloroform:isoamyl alcohol 24:1. The DNA was precipitated with 100% ethanol, omitting the addition of 3M NaOAc. The precipitated, washed and dried pellet was re-suspended in either 50 µL MilliQ water or TE. The requirement for high molecular weight DNA required that at all steps the DNA preparation was treated with extreme care, all pipetting was carried out with wide-bore tips and any mixing was achieved by gentle inversion of the tube.

The quality and quantity of DNA was estimated through electrophoresis on a low percentage agarose gel (0.6%) and by scanning spectrophotometry between $A_{360} - A_{200}$.

3.2.2.2 Purification of extracted DNA by density gradient centrifugation

Prior to library construction with the λ DASH II vector, genomic DNA was purified on a caesium chloride density gradient. A large scale DNA extraction from 2000 *C. helgolandicus* individuals was performed following a modification of the method in Section 3.2.2.1. Animals were placed in aliquots of 50 in microcentrifuge tubes. Each aliquot was homogenised with the addition of 440 µL homogenising solution (400mM Tris-HCl pH 8, 60 mM EDTA, 150 mM NaCl, 1% SDS) and 11 µL proteinase K (10 mg mL^{-1}) and incubated at 37°C for 30 min. The homogenate was then treated as in Section 3.2.2.1 and re-suspended in 1 mL MilliQ water.

To the DNA solution 1.5 mL of TE buffer, 2.56 g caesium chloride and 0.25 mL ethidium bromide (10 mg mL^{-1}) were added. The prepared solution was transferred to two microcentrifuge tubes with a wide-bore 1 mL tip and centrifuged for 2 min (RCF=11600) to remove large debris. The clean sample was transferred using a wide bore needle (21G) into a 3 mL Beckman centrifuge tube, filling to half way up the neck, ensuring that no air bubbles were present. A metal cap was placed on top of the centrifuge tube and the tube sealed by heating the metal cap and applying pressure. The sample was centrifuged in a Beckman TL100 centrifuge (RCF = 100,000) for 18 h at 20°C.

The DNA band was viewed by UV transillumination and recovered via the insertion of a large bore needle through the side of the tube. The recovered DNA was repeatedly extracted with isopropanol (saturated with distilled water and caesium chloride) until all traces of ethidium bromide were removed. The caesium chloride saturated DNA solution was aliquoted into microcentrifuge tubes and precipitated with 2.5 volumes 100% ethanol (-20°C). The precipitated DNA was centrifuged for 2 min (RCF=11600), the

supernatant removed and the pellet washed repeatedly with 70% ethanol (room temperature) to remove caesium chloride. The DNA pellet was re-suspended in 750 µL MilliQ water at 37°C overnight.

The quantity and quality of DNA were viewed by agarose gel electrophoresis and the A_{260} A_{280} ratio measured. The purity of the DNA was also tested by performing small-scale restriction digests and a test self-ligation.

3.2.3 Partial restriction endonuclease digestion of genomic DNA

Extracted high molecular weight genomic DNA was digested with *Sau*3AI to provide insert DNA ends compatible with the *Bam*HI-digested arms of the λ DASH II vector system.

To optimise the restriction endonuclease digestion the following small-scale reactions were performed.

Dilution buffer	
10 x *Sau*3AI buffer	20 µL
1 mg mL^{-1} BSA	20 µL
MilliQ water	160 µL

DNA solution	
Genomic DNA	10 µg
10 x *Sau*3AI buffer	20 µL
1 mg mL^{-1} BSA	20 µL
MilliQ water	to a final volume of 200 µL

The following dilutions (Table 3.1) were prepared on ice using *Sau*3AI (10 U μL^{-1}).

Tube no.	Preparation	Units *Sau*3AI µg DNA^{-1}
1	2 µL *Sau*3AI (20 units) + 98 µL dilution buffer	1
2	10 µL tube 1 + 10 µL dilution buffer	0.5
3	10 µL tube 2 + 10 µL dilution buffer	0.25
4	10 µL tube 3 + 10 µL dilution buffer	0.125
5	10 µL tube 4 + 10 µL dilution buffer	0.063
5	10 µL tube 5 + 10 µL dilution buffer	0.032
7	10 µL tube 6 + 10 µL dilution buffer	0.016
8	10 µL tube 7 + 10 µL dilution buffer	0.008

Table 3.1 Dilution series for *Sau*3AI partial digestion of genomic DNA.

20 µL aliquots of DNA solution were placed into separate labelled tubes. To each tube 5 µL of the appropriate enzyme dilution were added and mixed gently by pipetting. The restriction digests were then incubated at 37°C for 15 min followed by 10 min at 65°C to heat inactivate the enzyme. For each tube 12 µL of sample with 3 µL 5 x orange G loading buffer (40% w/v sucrose in water, 0.25% Orange G) were loaded on to a 0.4 % agarose gel along with 200 ng of λ *Hin*dIII DNA marker and 200 ng of λ *Sma*I DNA marker. Electrophoresis was performed at 2 volts/cm overnight.

The gel was viewed on an UV transilluminator to determine the amount of enzyme required to produce the maximum intensity of fluorescence in the desired size range.

With reference to the results shown in Section 3.3.3, two aliquots of 10 µg genomic DNA were digested with 0.065 and 0.031 units *Sau*3AI µg DNA^{-1}. Following incubation at 37°C for 15 min and heat inactivation at 65°C for 10 min the partially digested DNA was extracted once with phenol/chloroform:isoamyl alcohol 24:1 and once with

chloroform:isoamyl alcohol 24:1. The DNA was then precipitated and the pellet dried and re-suspended in 1 μL MilliQ water to give an approximate concentration of 10 μg μL^{-1}.

3.2.4 Assembly and titration of intact recombinant λ DASH II bacteriophage

3.2.4.1 Ligation of partially digested genomic DNA into λ DASH II

Two ligation reactions were set up in a total volume of 5 μL (A) and 10 μL (B) (Table 3.2). The ligations were performed at a DNA concentration ≥ 0.2 μg μL^{-1}, to favour the construction of concatamers over circular DNA molecules. Ligation A contained a vector:insert weight ratio of 1:2.5 and ligation B had a vector:insert weight ratio of 1:5. Ligation reactions were set up on ice and incubated a 4°C overnight.

	+ ve control	- ve control	Ligation A	Ligation B
λ DASH II vector (1 μg)	1 μL	1 μL	1 μL	1 μL
Insert (1 μg μL^{-1})	1.2 μL*	0 μL	2.5 μL	5 μL
10 x ligase buffer	0.5 μL	0.5 μL	0.5 μL	1.0 μL
10 mM rATP	0.5 μL	0.5 μL	0.5 μL	0.5 μL
T_4 DNA ligase (20 U μL^{-1})	0.5 μL	0.5 μL	0.5 μL	0.5 μL
MilliQ water	1.3 μL	2.5 μL	0 μL	2 μL
Total volume	5 μL	5 μL	5 μL	10 μL

* The positive control insert was 0.3 μg pME/*Bam*HI (supplied with the λ DASH II vector kit)

Table 3.2 Initial ligation reactions of *Sau*3AI partially-restricted genomic DNA with *Bam*HI predigested λ DASH II vector arms.

3.2.4.2 Preparation of the host bacteria strain

The host strain XL1-Blue MRA (P2) was prepared for use with the λ DASH II vector, and the host strain VCS257 was prepared for use with the positive wild-type lambda DNA packaging control. Cells were revived from storage at –80°C and streaked onto a LB agar plate. The plates were inverted and grown overnight at 37°C. A single colony for each host strain was used to inoculate a 500 mL flask containing 50 mL LB media (supplemented with 0.5 mL 20% maltose and 0.5 mL 1 M $MgSO_4$). The cells were grown with shaking at 37°C for 4-6 hours until OD_{600} approached 1.0. The cells were pelleted by centrifugation in a Falcon 50 mL BlueMax tube for 10 min (RCF = 500) at 4°C and gently re-suspended in half the original volume with sterile 10 mM $MgSO_4$. Directly before use the cells were diluted to an OD_{600} = 0.5 with sterile 10 mM $MgSO_4$.

3.2.4.3 Packaging and titration of recombinant λ DASH II

Gigapack®III Gold packaging extracts (Stratagene) were used to package the recombinant λ phage. The single-tube format of Gigapack III packaging extracts contains phage infected *E. coli* cell extract to supply the mixture of proteins needed for encapsidating λ DNA in one tube. The recombinant λ DNA was packaged as follows. The appropriate number of packaging extracts were removed from storage at –80°C and placed on ice. Each extract was thawed quickly and experimental recombinant DNA was added immediately. For the highest packaging efficiency, 1 μL of each recombinant DNA sample (ligation A, ligation B, +ve and -ve ligation controls) was added to each packaging extract. The tubes were then stirred with a pipette tip, ensuring that no air bubbles were introduced and the tube centrifuged (RCF = 11600) for approximately 5 seconds to ensure that all contents were at the bottom of the tube. The recombinant DNA and packaging extract were then incubated at room temperature for 2 h. Following the incubation 500 μL

SM buffer (100mM NaCl, 8 mM $MgSO_4$, 20 mM Tris-HCl pH 7.5, 0.01% w/v gelatin) were added. A further 20 µL of chloroform were added and the contents of the tube mixed gently and centrifuged briefly to sediment the debris. The supernatant containing the phage was stored at 4°C until use.

1:10 dilutions of the final packaged reactions (ligation A, ligation B, +ve and -ve ligation controls) were made in SM buffer (100 mM NaCl, 8 mM $MgSO_4.7H_2O$, 2% w/v gelatin). 1µL of undiluted and 1µL of the 1:10 dilution of the final packaging reactions were added to aliquots of 200 µL of prepared host cells diluted in 10 mM $MgSO_4$ to an OD_{600} of 0.5. The phage and bacteria were incubated at 37°C for 15 min to allow the phage to attach to the cells. Following incubation the infected bacteria were added to 3 mL of LB top agar (bactotryptone 10 g L^{-1}, NaCl 5 g L^{-1}, agarose 6 g L^{-1}), cooled to approximately 50°C. The LB top agar was gently mixed and then plated immediately onto LB agar bottom plates pre-warmed to 37°C. Once dry the plates were inverted and incubated at 37°C overnight.

The efficiency of the Gigapack III packaging extract was tested with the wild-type λ control DNA. The wild-type λ control DNA (approximately 0.2 µg of λc1857) was packaged following the methods above. 1 mL of SM buffer was added and the packaging reaction diluted 1:100 and 1:10,000 in SM buffer. 10 µL of each positive packaging control dilution were added to 200 µL of prepared VCS257 host cells.

Following overnight incubation the plaques on each plate were counted and the titre in plaque-forming units per millilitre (pfu mL^{-1}) determined as follows:

$$\frac{\text{(number of plaques) x (dilution factor)}}{\text{(volume of extract plated)}} = \text{pfu mL}^{-1}$$

The packaging efficiency of the arms was then calculated as follows:

$$\frac{\text{(pfu mL}^{-1}\text{)}}{\text{(concentration of vector DNA packaged)}} = \text{recombinants µg DNA}^{-1}$$

3.2.5 Dephosphorylation and size fractionation of partially digested DNA

An attempt to increase the titre of the library involved dephosphorylating and size fractionating the partially digested DNA prior to ligation. The DNA was prepared by purification on a caesium chloride density gradient and partial digestion with *Sau*3AI, following the optimised conditions determined in the *Sau*3AI partial digest assay (Section 3.3.3).

3.2.5.1 Dephosphorylation and Size fractionation on a sucrose gradient

In preparation for the sucrose gradient the following sucrose solutions were made and autoclaved:

10% sucrose solution	40% sucrose solution
10% sucrose	40 % sucrose
20 mM Tris-HCl pH 8	20 mM Tris-HCl pH 8
5 mM EDTA pH 8	5 mM EDTA pH 8

The sucrose gradients were constructed using a gradient former in conjunction with a peristaltic pump. The sucrose gradient was layered into a dome topped 14 mL Beckman centrifuge tube until the tube was filled to the turn of the neck. Once formed 110 µL of the sucrose gradient were removed and replaced with 110 µL of dephosphorylated DNA.

The DNA was dephosphorylated by combining 86 µL DNA (approximately 80 µg) with 11 µL 10 x buffer and 7 units of calf intestinal alkaline phosphatase (CIAP). The reagents were mixed gently and incubated for 30 min at 37°C. Following this initial incubation a further 5 units of CIAP were added and the incubation continued at 37°C for a further 30 min. Following the dephosphorylation the sample was stored briefly on ice and then heated to 65°C for 10 min. The DNA was loaded directly on to the sucrose gradient, the tube heat sealed and centrifuged in a Beckman L7-55 ultracentrifuge with a fixed arm rotor (50.2ti) at 22°C for 20 h (RCF = 50000).

The DNA was recovered from the sucrose gradient by carefully removing the tube from the centrifuge and placing it in a retort stand and clamp. The top of the tube was punctured with an 18G needle to create an air inlet and a second needle of the same size, connected to a length of tubing running through a peristaltic pump, was inserted through the bottom of the tube. Using the pump to control the flow, 30 aliquots of approximately 8 drops of sucrose gradient (approximately 50μL) were collected in microcentrifuge tubes. 20 μL of every other fraction were taken, mixed with 1 μL concentrated orange G and loaded onto a 0.5 % agarose gel with λ DNA *Hin*dIII and λ DNA *Sma*I size markers. Electrophoresis was performed at 8 volts/cm, and the gel viewed and photographed on an UV transilluminator to determine the fractions containing the desired DNA size range.

Fractions containing DNA of the required size were divided into 200 μL aliquots. Each aliquot of DNA in sucrose was precipitated by adding 1.5 volumes of TE, 0.2 volumes 3M NaOAc and 2.5 volumes 100% ethanol. The tubes were incubated at –20°C overnight and then -80°C for 8 hours. The precipitated DNA was centrifuged for 15 min (RCF=11600), and the pellets washed with 70% ethanol, dried and re-suspended in 2 μL MilliQ water. The aliquots were combined and 1μL of sample, plus 1 μL 2 x Orange G loading buffer, separated on a 0.5% gel. λ *Hin*dIII markers and standards of known concentrations of λ DNA were electrophoresed on the same gel to enable determination of size and concentration.

3.2.5.2 Ligation, packaging and titration

Ligation reactions were set up as shown in Table 3.3 with a vector:insert weight ratio of 1:0.625 for ligation 1 and 1:1.625 for ligation 2. This approximately equated to a molar ratio of vector:insert of 1:2 for ligation 1 and 1:5 for ligation 2. As before DNA

concentrations in the ligation were kept above 0.2 μg μL^{-1} to favour the production of concatamers. Ligation reactions were set up on ice and incubated a 4°C overnight.

	+ ve control	- ve control	Ligation 1	Ligation 2
λ DASH II vector (1 μg)	1 μL	1 μL	1 μL	1 μL
Insert (0.25 μg μL^{-1})	1.2 μL*	0 μL	2.5 μL	6.5 μL
10 x ligase buffer	0.5 μL	0.5 μL	0.5 μL	1.0 μL
10 mM rATP	0.5 μL	0.5 μL	0.5 μL	1.0 μL
T_4 DNA ligase (20 U μL^{-1})	0.5 μL	0.5 μL	0.5 μL	0.5 μL
MilliQ water	1.3 μL	2.5 μL	0 μL	0 μL
Total volume	5 μL	5 μL	5 μL	10 μL

* The positive control insert was 0.3 μg pME/*Bam*HI (supplied with the λ DASH II vector kit)

Table 3.3 Ligation of *Sau*3AI partially restricted genomic DNA (following dephosphorylation and size fractionation) with *Bam*HI predigested λ DASH II vector arms.

The host strains XLI-blue MRA (P2) and VCS257 were prepared as in Section 3.2.4.2. The ligations were packaged by adding 1 μL of each control ligation and ligation 1, and 2 μL of ligation 2 to 25 μL aliquots of thawed packaging extracts (Section 3.2.4.3). 500 μL of SM buffer were added to each tube, and 1 mL SM buffer to the packaged wild-type λ control DNA. 20 μL chloroform were added to each sample, the contents of the tubes gently mixed, and each tube centrifuged briefly to sediment any debris. The packaging reactions were titrated with two consecutive 10^{-2} dilutions in SM buffer for the positive packaging control, and a 1:10 dilution in SM buffer for each of the remaining ligations. For the packaging control 10 μL of the 1:100 and 1: 10,000 dilutions were added to 200 μL of prepared host strain VCS527. For the remaining packaging reactions 1 μL of

undiluted and 1 µL of the 1:10 dilutions were added to 200 µL of XLI-blue MRA (P2) prepared host strain. The phage and bacteria mixes were incubated at 37°C for 15 min, added to 3 mL of TB top agar (50°C), and spread immediately onto pre-warmed LB bottom agar plates. The plates were allowed to dry and then incubated inverted overnight at 37°C.

Following overnight incubation the plaques on each plate were counted and the titre in plaque-forming units per mL (pfu mL^{-1}), and the packaging efficiency of the arms as recombinants per microgram DNA determined as in Section 3.2.4.3

3.2.6 Library amplification

Using the following equation of Clarke and Carbon, (1976) the potential for non-representation of a given sequence was minimised during library amplification.

$$N = \frac{\ln (1 - P)}{\ln (1 - \frac{1}{n})}$$

where: N = the number of independent recombinants

P = probability of the presence of any given sequence in the genomic library

n = size of the genome relative to a single cloned fragment

The haploid genome mass for *C. helgolandicus* is 10.5 pg/nucleus (McLaren *et al.*, 1988). As 1 pg = 9.1 x 10^8 bp the size of *C. helgolandicus* genome is therefore 9.55 x 10^9 bp.

The size of the genome relative to a single cloned fragment (n) is:

$$n = \frac{9.55 \times 10^9}{2.0 \times 10^4} = 4.78 \times 10^5$$

Therefore with a 95% probability of including any given sequence in the *C. helgolandicus* genome the number of independent recombinants (N) is 1.43 x 10^6.

$$N = \frac{\ln (1 - 95)}{\ln (1 - \frac{1}{4.78 \times 10^5})}$$

To increase the number of packaged recombinants to the required amount for a complete representation of the genome, a further 2 µL of ligation 1, and 3 x 2 µL aliquots of ligation 2 (Section 3.2.5.2) were packaged into the λ protein coats and infected into the host strain as described above (Section 3.2.5.2).

The 8 packaged mixtures or library suspensions of ligations 1 and 2 were divided into aliquots each containing approximately 5 x 10^4 pfu of bacteriophage, taking into consideration the background of non-recombinants as calculated by the negative ligation control. The 30 aliquots containing approximately 5 x 10^4 pfu of bacteriophage were sufficient to allow amplification of a total of 1.5 x 10^6 recombinants.

To amplify the library the host strain XL1-Blue MRA (P2) was prepared as in Section 3.2.4.2. The cells were diluted to an OD_{600} of 0.5 in 10mM sterile $MgSO_4$. The aliquots of packaged mixture containing approximately 5 x 10^4 pfu of bacteriophage were combined with 600 µL of host cells at an OD_{600} of 0.5 in Falcon 2059 polypropylene tubes. To amplify 1.43 x 10^6 plaques, a total of 30 aliquots were used (each aliquot containing 5 x 10^4 plaques 140 mm $plate^{-1}$). The tubes containing the phage and bacteria cells were incubated at 37°C for 15 min. Each aliquot of infected bacteria was mixed with 6.5 mL TB top agar cooled to 50°C, and spread evenly onto a freshly poured 140 mm bottom agar plate. The plates were allowed to dry and incubated inverted at 37°C for 7 h 15 min until the plaques were no larger than 1-2 mm and were just touching. The plates were over-laid with 10 mL SM buffer, and incubated at 4°C with gentle rocking overnight.

The following day the bacteriophage suspension was recovered from each plate with a 5 mL pipette and the suspensions pooled in sterile Falcon 50 mL BlueMax tubes. The

plates were rinsed with an additional 1.5 mL SM buffer and added to the pool of suspension. Chloroform was added to a 5% (v/v) final concentration and the suspension mixed well and incubated for 15 min at room temperature. The cell debris was removed by centrifugation for 10 min (RCF = 500). The supernatant was transferred to a clean Falcon 50 mL BlueMax tube and the above step repeated to ensure removal of all cellular debris. The supernatant was again transferred to a clean Falcon tube. To an aliquot of the library, chloroform was added to a final concentration of 0.3% (v/v) and the supernatant stored at 4°C. Further aliquots of the library were stored at –80°C with the addition of 7% (v/v) DMSO.

The titre of the amplified library was checked using host cells and serial dilutions of the library.

3.3 Results

3.3.1 Overview of library construction

Cloning of *C. helgolandicus* genomic DNA into the replacement vector λ GEM-12 repeatedly proved unsuccessful. Increasing the purity of the sample DNA had no positive effect and the addition of a small quantity of ligated *C. helgolandicus* DNA to both the positive ligation control and the positive transformation control led to no reduction in titre.

Further factors were tested to determine the cause of the problem but no alterations to the protocol would yield plaques of recombinant phages. The results of test ligations showed that digested genomic DNA and digested plasmid DNA could be correctly self-ligated. However, both genomic and plasmid digested/filled-in DNA, was erroneously self-ligated.

3.3.2 DNA extraction from *Calanus helgolandicus*

3.3.2.1 Basic extraction procedure

The quality and quantity of extracted DNA was measured by visualisation on a 0.6% agarose gel (Figure 3.1) and spectrophotometrically (Figure 3.2). The concentration of DNA was estimated in μg mL^{-1} according to the equation:-

DNA concentration (μg mL^{-1}) = $(A_{260})(50)$(dilution factor)

The purity of DNA was calculated by dividing the A_{260} by A_{280}, the optimum value for pure double stranded DNA being 1.8. The concentration of the genomic DNA was 0.44 μg mL^{-1}, and the purity was 1.9.

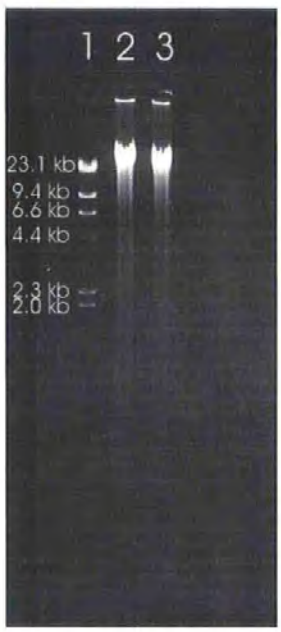


Figure 3.1 Agarose gel (0.6%) of high molecular weight *Calanus helgolandicus* genomic DNA. Lane 1 = 200 ng λ *Hin*dIII marker; Lane 2 + 3 = replicates of 1 μL of high molecular weight *Calanus helgolandicus* genomic DNA.

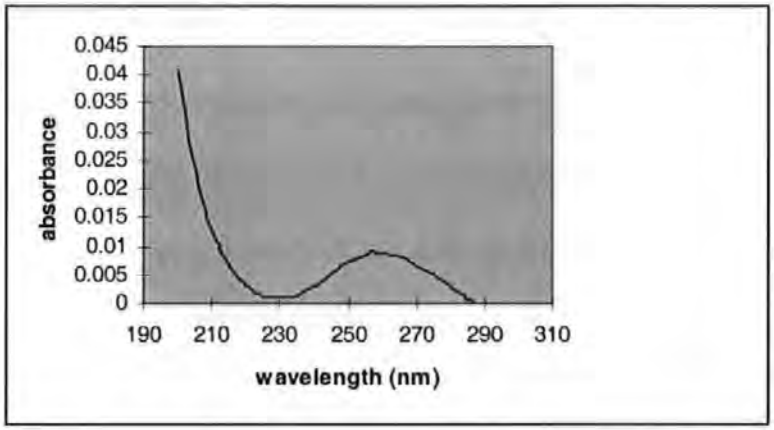


Figure 3.2 Spectrophotometric analysis ($A_{360} - A_{200}$) of 1 µL high molecular weight *Calanus helgolandicus* genomic DNA in 999 µL MilliQ water. A_{260} = 0.0088, A_{280} = 0.0046. Concentration of DNA = 0.44 µg mL^{-1}, purity = 1.9.

3.3.2.2 Purification of extracted DNA by density gradient centrifugation

The quantity and quality of the genomic DNA following purification on a caesium chloride gradient was determined by visualisation on a 0.6% agarose gel (Figure 3.3) and spectrophotometrically (Figure 3.4). The concentration of the DNA was approximately 1.0 µg mL^{-1} and the purity = 1.8.

Figure 3.3 Agarose gel (0.6%) of high molecular weight *Calanus helgolandicus* genomic DNA following purification on a caesium chloride gradient. Lane 1 = 200 ng λ *Hin*dIII marker; Lane 2 = 2 µL of purified genomic DNA.

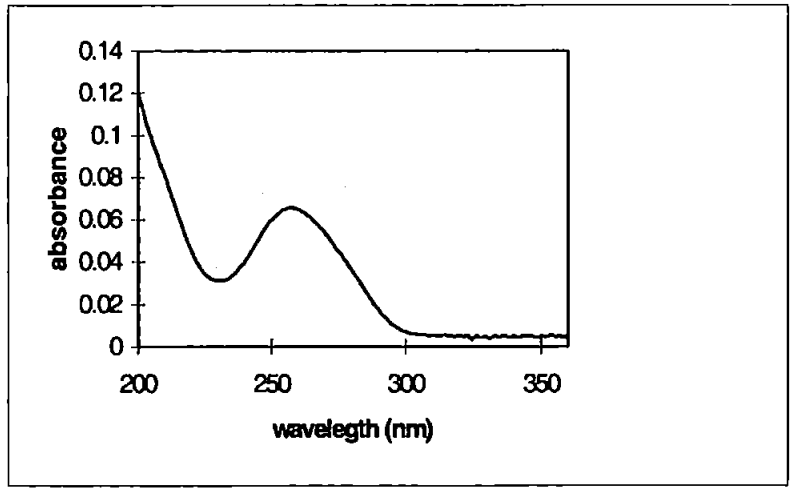


Figure 3.4 Spectrophotometric analysis ($A_{360} - A_{200}$) of 3 µL high molecular weight *Calanus helgolandicus* genomic DNA following purification on a caesium chloride gradient, in 997 µL MilliQ water. $A_{260} = 0.0647$, $A_{280} = 0.0359$. Concentration of DNA = approximately 1.0 µg mL^{-1}, purity = 1.8.

3.3.3 Partial restriction endonuclease digestion of genomic DNA

The results of the small-scale optimisation of *Sau*3AI partial digestion of genomic DNA can be seen in Figure 3.5. The reactions were separated on an agarose gel (0.4%) and the gel viewed by UV transillumination to determine the amount of enzyme required to produce the maximum number of molecules in the desired size range.

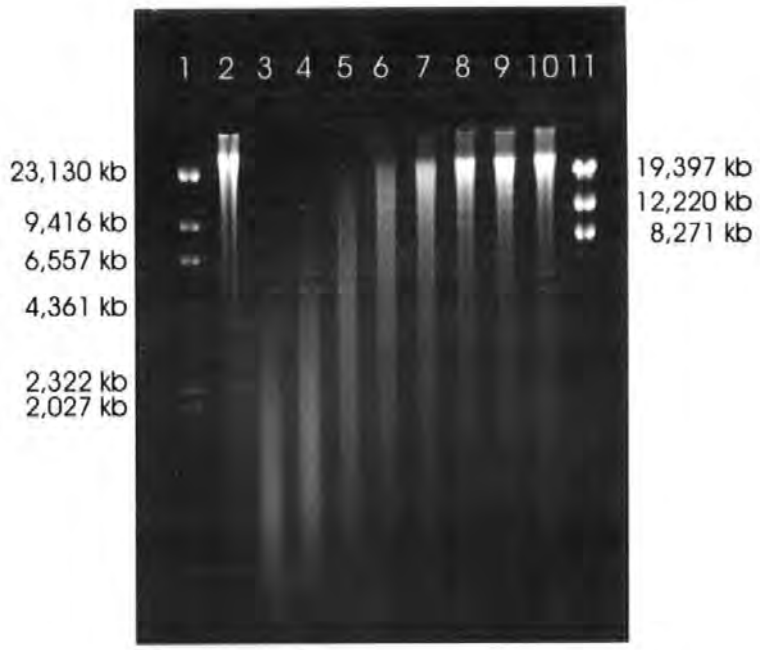


Figure 3.5 Agarose gel (0.4%) showing assay of *Sau*3AI partial digestion of genomic DNA. Lane 1 = 200 ng λ *Hin*dIII marker; lane 2 = approximately 0.5 μg undigested genomic DNA; lanes 3-10 = 12 μL of digestion reactions tube numbers 1-8 respectively. Lane 11 = 200 ng λ *Sma*I marker.

Reactions 4 and 5, with an enzyme concentration of 0.125 and 0.063 units μg DNA^{-1} respectively, showed the highest amount of fluorescence just below 20 kb. Therefore half this amount of enzyme μg DNA^{-1} was used for two large scale restriction digests (Section 3.2.1.1).

3.3.4 Assembly and titration of intact recombinant bacteriophage

The results from the initial library constructed without dephosphorylation and size fractionation of the *C. helgolandicus* insert DNA are shown in Table 3.4.

Sample		Number of plaques plate^{-1}	Titre of phage pfu mL^{-1}	Packaging efficiency of arms Recombinants µg DNA^{-1}
+ve packaging control	1:100 dilution	excess	-	-
	1:10,000 dilution	704	7.04×10^8	3.52×10^9
+ve ligation control	undiluted	740	7.4×10^5	1.48×10^6
	1:10 dilution	72	7.2×10^5	1.44×10^6
-ve ligation control	undiluted	9	9×10^3	2.37×10^4
	1:10 dilution	0	-	-
Ligation A	undiluted	208	2.08×10^5	1.56×10^5
	1:10 dilution	21	2.15×10^5	1.58×10^5
Ligation B	undiluted	107	1.07×10^5	4.67×10^4
	1:10 dilution	9	9×10^4	3.9×10^4

Table 3.4 Titre of phage (pfu mL^{-1}) and packaging efficiency of arms (recombinants µg DNA^{-1}) of the initial library.

The titre of the phage for the sample ligations A and B (Table 3.3) indicated that *C. helgolandicus* DNA had successfully been cloned into the λ DASH II vector arms. The positive packaging control indicated that the Gigapack III packaging extract was working to a high efficiency (uncut wild-type λ DNA is recommended to package with an efficiency exceeding 1 x 10^9 pfu μg^{-1}). The ligation control results indicated that pME/*Bam*HI inserts successfully ligated to the λ DASH II vector arms, and that there were acceptably low levels of non-recombinants. The positive ligation control however, could theoretically be increased by approximately a factor of 5, the optimal titre being between 5 x 10^6 – 1 x 10^7. For a complete representation of the *C. helgolandicus* genome the number of pfu mL^{-1} of the sample ligations are below their optimum.

3.3.5 Increase in titre with size fractionation and dephosphorylation of DNA inserts

To increase the titre of the library the partially restricted DNA was dephosphorylated and size fractionated prior to cloning. The results of the large-scale partial digest are shown in Figure 3.6.

Figure 3.6 Large-scale *Sau3A*I partial digest of *Calanus helgolandicus* genomic DNA. Lane 1 and 4 = 200 ng λ *Hin*dIII DNA marker; lane 2 = approximately 500 ng of CsCl purified, partially digested *Calanus helgolandicus* DNA; lane 3 = approximately 250 ng of CsCl purified, undigested *Calanus helgolandicus* DNA.

Following size fractionation on a sucrose gradient every second fraction was separated by electrophoresis to determine fractions containing the desired size range of DNA. The results are shown in Figure 3.7.

Figure 3.7 0.5% agarose gel of *Calanus helgolandicus* genomic DNA size-fractionated on a sucrose gradient. Lane 1 = λ *Hin*dIII marker; lanes 2-15 = 20 μL of sucrose gradient fractions.

The first 11 fractions (as indicated by lanes 2-7) contained DNA of the required size, with the majority of these fractions containing DNA just below 20 kb. The precipitated and re-suspended fractions were pooled. Figure 3.8 shows 1 μL of the size fractionated partially restricted pooled DNA.

Figure 3.8 Partially digested, dephosphorylated *Calanus helgolandicus* genomic DNA following size fractionation on a sucrose gradient. Lane 1 = 200 ng λ *Hin*dIII marker; lane 2 = 100 ng λ DNA standard; lane 3 = 200 ng λ DNA standard; lane 4 = 400 ng λ DNA standard; lane 5 = 1 μL digested, dephosphorylated, size-fractionated *Calanus helgolandicus* genomic DNA; lane 6 = 200 ng λ *Hin*dIII marker. Each sample was loaded with Orange G loading buffer.

The size of the DNA appeared to be between 9 and 23 kb, with the majority of the DNA being towards the higher of this range. With comparison to the λ standards the concentration was approximated to be 0.25 μg μL^{-1}.

The titre in plaque-forming units per mL (pfu mL^{-1}) and the packaging efficiency of the arms as recombinants per microgram of DNA were determined for the cloned dephosphorylated and size-fractionated *C. helgolandicus* DNA. The results are shown in Table 3.5.

Sample		Number of plaques plate $^{-1}$	Titre of phage pfu mL^{-1}	Packaging efficiency of arms Recombinants µg^{-1}
+ve packaging control	1:100 dilution	excess	-	-
	1:10,000 dilution	260	2.6×10^8	1.3×10^9
+ve ligation control	undiluted	excess	-	-
	1:10 dilution	216	2.16×10^6	4.32×10^6
-ve ligation control	undiluted	4	3×10^3	7.89×10^3
	1:10 dilution	0	-	-
Ligation 1	undiluted	500	5×10^5	8.20×10^5
	1:10 dilution	54	5.4×10^5	8.85×10^5
Ligation 2	undiluted	800	8×10^5	3.49×10^5
	1:10 dilution	90	9×10^5	3.93×10^5

Table 3.5 Results of the increased efficiency library as titre of phage (pfu mL^{-1}) and packaging efficiency of arms (recombinants µg DNA^{-1}).

Both the titre of phage and packaging efficiency of arms has increased for ligation 1 and ligation 2, compared to ligation A and ligation B, following dephosphorylation and size fractionation of the insert DNA. Ligation 2, of the increased efficiency cloning

reactions, yielded an increase of plaque forming units per packaging reaction, but fewer plaque-forming units per microgram of DNA, compared to ligation 1.

3.3.6 Library Amplification

In order to grow enough plaques for a true representation of the *C. helgolandicus* genome further packaging reactions were performed on ligations 1 and 2. The titre of the new packaging reactions was 5.5 x 10^5 pfu mL^{-1} for ligation 1 and an average titre of 3.4 x 10^5 pfu mL^{-1} for the ligation 2 aliquots.

A sufficient number of plaques could then be grown to produce a representative genomic library for *C. helgolandicus*. This library was successfully amplified to provide a large quantity of high titre stock. The titre of the amplified library was checked with host cells and serial dilutions of the library, the average titre being 8.2 x 10^8 pfu mL^{-1}.

3.4 Discussion

A representative library for *Calanus helgolandicus* has been established in the replacement vector λ DASH II. A sufficient number of independent recombinants, 1.43 x 10^6, provides a 95% probability that any given sequence in the *C. helgolandicus* genome will be represented. This library has been stored long term at –80°C, providing a resource for the cloning of any *C. helgolandicus* genes.

The choice of vector for library construction is of the utmost importance. Having selected a lambda replacement vector, the choice of which particular vector is still critical. In theory, the use of LambdaGEM-12 half-site arms as a cloning vector has a high potential. The prevention of self-ligation reactions of vector arms and genomic fragments should have been ideal for use with small amounts of starting material, as there is no need for size fractionation or alkaline phosphatase treatment. However, only after many tests and thorough checking of each step of the process, was it possible to conclude that the

partial fill-in reaction was ineffectual. Both lambda and genomic (digested and partially filled-in) DNA were easily self-ligated, strongly suggesting that the fill-in reactions were unsuccessful. Further tests would have necessitated the use of radioactively labelled nucleotides within the fill-in buffer and subsequent detection of incorporated labelled bases in the filled-in DNA. Such tests would be both expensive and time consuming, and although they may confirm the problem they would be unlikely to provide a solution. Hence it was decided to choose a different λ replacement vector.

The cloning vector, lambda DASH II, proved to be a more successful cloning system, as indicated by the results in Table 3.5. The high efficiency of the positive packaging control compared to the positive ligation control is explained by the fact that uncut wild-type lambda DNA packages with a higher efficiency than predigested vector. The *spi* selection used by the λ DASH II cloning system also ensured that the number of non-recombinants, as seen in the negative ligation reaction, were low. The two sample ligations A and B proved that *C. helgolandicus* DNA had successfully been cloned. The greater yield of plaque forming units for ligation A compared with ligation B suggests that a lower vector:insert weight ratio gave more optimal ligation conditions.

The cloning results indicated that with an increase in efficiency, a successful library could be constructed. The combination of dephosphorylation and size fractionation of the insert DNA, in addition to the use of a suitably low vector:insert ratio successfully increased the efficiency of the cloning reactions approximately 5 fold. The dephosphorylation and size fractionation also prevented the cloning of DNA fragments that are not contiguous in the genome.

Ligation 2, of the increased efficiency cloning reactions, yielded an increase of plaque forming units per packaging reaction, but fewer plaque-forming units per microgram of DNA, compared to ligation 1. This is because a greater amount of ligation reaction was packaged for ligation 2, which although increasing the titre of the phage, does

not increase the packaging efficiency of the arms. The increase in efficiency of these cloning reactions was sufficient to construct successfully a genomic library for *C. helgolandicus.* The amplification of the library has enabled a large, stable quantity of high-titre stock to be stored.

The production of this genomic library provides a significant resource for any investigation involving the molecular genetics of *C. helgolandicus.* Specifically, it provides an excellent foundation from which to obtain *C. helgolandicus* developmental genes. The mechanisms by which putative developmental genes were obtained from this library are explained in Chapter4.

CHAPTER 4

Identification and analysis of specific developmental genes from the genomic library

CHAPTER 4

Identification and analysis of specific developmental genes from the genomic library

4.1 Introduction

4.1.1 Probe selection

4.1.1.1 Heterologous probes

4.1.1.2 Homologous probes

4.2 Methods

4.2.1 Obtaining probes to screen the library

4.2.1.1 Heterologous probes

4.2.1.2 Homologous probes

4.2.2 Screening the library

4.2.2.1 Choice of label

4.2.2.2 Complete representation of Calanus helgolandicus *genome*

4.2.2.3 Plating out the library for screening

4.2.2.4 Transfer of plaques to Hybond-N membranes

4.2.2.5 Proteinase K treatment of membranes

4.2.2.6 Preparation of ^{32}P-labelled probes

4.2.2.7 Hybridisation of probes to Hybond-N membranes

4.2.2.8 Removal of probes from Hybond-N membranes

4.2.2.9 Picking the positive plaques

4.2.2.10 Amplification of the positive plaques

4.2.2.11 Primary library screens

4.3 Results

4.1 Introduction

Embryogenesis plays a critical role in the life cycle of *Calanus*, yet the molecular basis of this developmental phase is not understood. The identification and analysis of developmental genes from *Calanus* will open new avenues in which to explore the reproductive and developmental biology of this important member of the zooplankton community. An understanding of the expression of *Calanus* developmental genes would enable the investigation of the molecular basis of environmental influence on development. The acquisition of such genes may also be applied to the improvement of existing techniques for the key measurement of *Calanus* population dynamics, such as *Calanus* egg viability. A suitable *Calanus* developmental gene, expressed in early embryogenesis, would allow the development of a molecular system with which to ascertain the developmental condition of *Calanus* eggs.

The objective of Chapter 4 is to identify and analyse cloned fragments of *C. helgolandicus* developmental genes by probing the *C. helgolandicus* library (Chapter 3). An important factor is the choice, construction and labelling of the probes for the screening of the library.

4.1.1 Probe selection

4.1.1.1 Heterologous probes

Many genes have been sufficiently conserved in evolution such that cross-species nucleic acid hybridisation is possible (Schuh *et al*, 1986). One potential approach to library screening utilises such cross species hybridisation, whereby gene sequences used to design the probes originate from different organisms. Such probes are termed "heterologous gene probes". This system enables specific genes to be targeted, based on previous characterisation of the gene in the organism from which the sequence originates.

The method is directly applicable to the cloning of developmental genes, whereby the homeobox and flanking elements constitute viable heterologous probes (Smerdon, 1998).

A prime candidate for a source of characterised homologous developmental genes is the fruitfly, *Drosophila melanogaster,* this organism having been extensively characterised with respect to developmental mechanisms. Interest in genes with potential use as an egg viability probe in *Calanus* lead to the selection of genes that were expressed during early embryogenesis in *Drosophila.* Such *Drosophila* genes would be suitable for use as heterologous probes.

One fundamental principle of developmental biology is that growth and differentiation result from genes being turned on and off, resulting in a so-called developmental cascade. In *Drosophila* this developmental cascade begins with the co-ordinate genes. These are first expressed maternally during oogenesis and determine the anteroposterior axis (Nüsslein-Volhard, 1991). Secondly there are the segmentation genes, which lead to the conversion of the gradient patterns into a periodic pattern of body segments. These segmentation genes are divided into three groups. The first group, the gap genes, affects regions several segments wide. The second group of segmentation genes, the pair-rule genes, affects alternate segments. The third group, polarity genes, incorporates genes that act in every segment. Finally there are the homeotic genes; these being master genes that convert the repetitive pattern of segments into individual segments, each with its own identity.

For the purpose of screening the library the co-ordinate genes were not considered, as such maternally expressed genes would not necessarily be indicative of embryonic development and therefore, not suitable for use an egg viability probe. The main class of developmental genes considered were the segmentation genes, in particular the gap-class genes, as these are expressed post-fertilisation and early in embryogenesis in *Drososphila.* The gap-class genes chosen to screen the library included *tailless* (Pignoni *et al.*, 1990),

giant (Capovilla *et al.*, 1992), *kruppel* (Schuh *et al.*, 1986), and *orthodenticle* (Finkelstein *et al.*, 1990). Genes from the segmentation pair-rule class include *hairy* (Feder *et al.*, 1994; Rushlow *et al.*, 1989), *fushi-tarazu, runt* (Kania *et al.*, 1990), and *even-skipped* (Macdonald *et al.*, 1986). The sequences of these genes, acquired from the computer database, were used to design primers with which to amplify the gene of interest. Homologous sequences for each gene were also acquired from the database and aligned with the *Drosophlia* gene. The sequences for each gene were then scrutinised to find conserved regions. Wherever possible the primers were designed to these conserved regions, in the hope that they might also prove suitable for the cross-species amplification of the gene from *C. helgolandicus* genomic DNA. Unfortunately, conserved regions were very limited (with the exception of the homeobox, Section 4.1.1.2) and no primers were able to amplify successfully developmental genes from *C. helgolandicus* genomic DNA. The primers were however, successful in the amplification of the *Drosophila* genes that were then used as heterologous probes with which to screen the library.

4.1.1.2 Homologous probes

When gene homologs have been cloned and sequenced from a range of different organisms, it is often possible to identify regions of conserved sequence identity from computer generated sequence alignments. If this gene identity exists over a moderate length of DNA, it is possible to design degenerate oligonucleotides to each end of the single conserved region in order to amplify the equivalent portion of gene from other organisms. If amplified from the organism under study, the conserved gene fragment can be used as a homologous probe to screen a library from that particular organism in order to obtain a fragment of the gene containing sequence data outside the conserved region.

The strategy of degenerate oligonucleotide amplification has been used extensively to amplify a relatively small conserved region of developmental genes known as the

homeobox. Using this technique, homeoboxes have been found in many organisms including vertebrates (Graham *et al.*, 1989), nematodes (Bürglin *et al.*, 1991), sponges (Seimiya *et al.*, 1994; Degnan *et al.*, 1995), and cnidarians (Schierwater *et al.*, 1991).

The homeobox is a DNA sequence, approximately 180 bp in length, common to many developmental genes. It codes for a highly conserved stretch of approximately 60 amino acids known as the homeodomain. The homeodomain is highly conserved through evolution and plays a major role in developmental regulation. The consensus sequence of the homeodomain (Bürglin, 1994) is shown in Figure 4.1. The protein sequences of helix I and helix III, indicated in Figure 4.1 are the most conserved, and have therefore been used to design 'forward' and 'reverse' degenerate oligonucleotide primers. The oligonucleotide primers are designed from reverse translation of the least degenerate protein regions (Smerdon, 1998). These primers, shown in Figure 4.2, enable a region of the homeodomain to be amplified using the polymerase chain reaction.

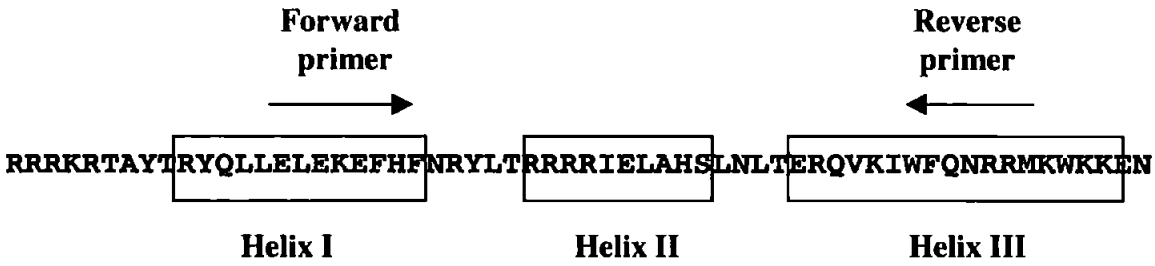


Figure 4.1 Homeodomain amino acid concensus sequence based on 346 sequences (Bürglin,1994). The three α helices are boxed. The regions of the forward and reverse oligonucleotide primers are indicated with an arrow above the consensus sequence.

Forward Primer – 'Hox C'

```
                AC     A  A  A  C
5' - GCTCAGA--TNGA-AA-GA-TT-CA - 3'
                GT     G  G  G  T
```

Reverse Primer – 'Hox D'

```
                  G  GG  C  G
5' - GGAATTCATNC-NC--TT-TG-AACCA - 3'
                  T  TA  T  A
```

Figure 4.2 Degenerate oligonucleotide primers designed for the amplification of homeobox sequences.

Using these degenerate primers homeoboxes have been found to exist in *C. helgolandicus* (Smerdon, 1998). A number of these *C. helgolandicus* homeoboxes have been cloned and sequenced (Smerdon, unpublished), and two such homeoboxes were selected for use as homologous probes to screen the library, based on predicted function from the amino acid sequences. Utilising homologous homeobox sequences to probe the library is specific only for homeobox genes (specificity to the level of identical homeobox sequences is very difficult to achieve), and therefore, considering the conserved nature of the homeodomain, would be expected to target a number of homeobox-containing genes. The expression profiles of such genes would then require examination in order to assess their suitability as an egg viability probe.

4.2 Methods

4.2.1 Obtaining probes to screen the library

4.2.1.1 Heterologous probes

Homologous developmental genes were obtained by designing primers to *Drosophila melanogaster* gene sequences obtained from the EMBL computer database. These primers were then used to amplify a region of the gene of interest, from *Drosophila* genomic DNA, using the PCR. The following primers were designed from *Drosophila* sequence, but account was taken of regions that were conserved with other organisms, such that the primers had potential use for direct amplification of the equivalent gene from *C. helgolandicus* genomic DNA.

Segmentation gap-class genes	**Forward primer**	**Reverse primer**
Tailless	5'-cattacggcatctacgcc-3'	5'-gttggttcagcacctcc-3'
Giant	5'-gttcagggatccttgg-3'	5'-cacttcttgagttggtcc-3'
Kruppel	5'-cacccacactggtgaac-3'	5'-ggactggatgaatgtcttc-3'
Orthodenticle	5'-tatccaggcgtcaacacacg-3'	5'-actccggcaccaatatcacc-3'

Segmentation pair rule genes	**Forward primer**	**Reverse primer**
Hairy	5'-gaagacagtaaagcatctgc-3'	5'-ctcctcctccttgatctg-3'
Fushi-tarazu	5'-taccacaaacagccagagcc-3'	5'-ctgatgccaaagtctcctcg-3'
Runt	5'-atggctatccccatcccg-3'	5'-cttgtccaggtcgaactcc-3'
Even-skipped	5'-ggataactccttgaacggc-3'	5'-attgagccactggactgc-3'

The components of each PCR reaction were as follows:

dNTPs	-5 μL
10x Dynazyme buffer	-5 μL
100 ng.μL^{-1} Forward primer	-5 μL
100 ng.μL^{-1} Reverse primer	-5 μL
Dynazyme	-1 μL
MilliQ water	-28 μL
Drosophila genomic DNA (300ng.μL^{-1})	-1 μL

Amplification was performed in an Autogene thermocycler (Grant). The cycling parameters included an initial denaturation step of 95°C (5min), followed by 35 cycles of specific Tm (1 min), 72°C (2 min), and 95°C (1 min). A final annealing phase at the specific Tm (2 min) was followed by an extension phase at 72°C (5 min) and storage at 4°C. The specific Tm found to be optimum for each primer pair was as follows; *kruppel* 60°C; *hairy* and *giant* 62°C; *tailless*, *runt* and *fushi-tarazu*, 65°C; *even-skipped* and *orthodenticle*, 68°C.

Amplification efficiencies were checked by 1% agarose gel electrophoresis. To confirm product identity the amplified fragments were then cloned and sequenced. Amplifications were prepared for cloning by excision from 1% agarose gel followed by recovery of the DNA using a Qiaex II gel extraction kit (Qiagen). The gene fragments were cloned in to pBluescript following the protocols detailed in Sections 2.3.8-2.3.10, and 2.2.5. Recombinant clones were identified by colony PCR (Section 2.3.11), and sequenced by the chain-termination method (Sanger *et al.*, 1977), following the protocols detailed in

Sections 2.3.12-2.3.17. The sequence data was used to confirm the identity of the amplified products by comparison with the original *Drosophila* sequence.

To prepare gene fragments for labelling the recombinant plasmid DNA was recovered from *E. coli*, and the inserts removed by endonuclease restriction. The initial probes prepared for labelling were *tailless, orthodenticle, hairy* and *even-skipped.* Approximately 4 μg of plasmid preparation for each fragment were digested with 10 units of *Eco*RI and *Hin*dIII in 20 μL of reaction mix containing 2 μL multicore 10 x buffer and 2 μL 1 mg mL^{-1} BSA. The reactions were incubated at 37°C overnight. The complete digest reaction was separated by 1.2% gel electrophoresis and the bands corresponding to the correct size inserts excised. The DNA was recovered from the agarose gel using a Qiaex II gel extraction kit (Qiagen), and stored ready for labelling.

4.2.1.2 Homologous probes

A number of *C. helgolandicus* homeoboxes have been previously amplified, cloned in pPCR-script and sequenced (Smerdon, 1998). Two such clones, referred to as clone 207 and clone 228, were prepared for labelling by amplifying the inserted homeobox sequences with the degenerate homeobox primers Hox C and Hox D. The amplified products were separated by 1% agarose gel electrophoresis, recovered from the gel using a Qiaex II gel extraction kit (Qiagen), and stored ready for labelling.

4.2.2 Screening the library

4.2.2.1 Choice of label

A suitable label with which the probes could be marked and subsequently used to probe the library was investigated. Radioactive labelling methods are certainly the most traditional. However, because of the hazards associated with radioisotope usage, the first choice was to label the hybridisation probe in a non-radioactive manner. The system

chosen was commercially available from Boehringer Mannheim and involves dUTP nucleotides modified by reaction with digoxigenin (DIG), a steroid hapten, to label DNA for hybridisation. The hybridised probes are immunodetected with anti-digoxigenin conjugated to alkaline phosphatase and then observed with the chemiluminescence substrate CSPD®. Enzymatic dephosphorylation of CSPD® by alkaline phosphatase leads to a light emission that is recorded on X-ray films. Although the apparent potential for the use of such a non-radioactive labelling technique is high, the DIG chemiluminescent kit proved completely inadequate and unsuitable for screening a large number of plaques. Instead, the more traditional method of using a labelled Phosphorous^{32} dATP was used.

4.2.2.2 Complete representation of the Calanus helgolandicus *genome*

In order to screen the library, a sufficient number of clones to represent the complete genome of *C. helgolandicus* have to be grown and transferred to a membrane. Clarke and Carbons equation (1976) (Section 3.2.6) was used to determine the number of clones needed to give a 95% probability of the library containing any particular gene. Taking into account the size of the *C. helgolandicus* genome, the size of the genome relative to a single cloned fragment, and with a 95% probability of including any given sequence in the *C. helgolandicus* genome, the number of independent recombinants needed was 1.43×10^6. With the maximum number of plaques that can be grown on a large LB agar plate being 5×10^4, a total of 30 plates was needed to give a 95% probability of any given sequence being represented.

4.2.2.3 Plating out the library for screening

Before the library was plated out the titre of the amplified and stored bacteriophage was checked. This was achieved by preparing the host strain XL1-Blue MRA (P2) (Section 3.2.4.2). Serial dilutions of the bacteriophage, ranging from a 1 in 10 dilution to a

1 in 100,000 dilution, were made in SM buffer (100 mM NaCl, 8 mM $MgSO_4.7H_2O$ and 2% w/v gelatin). A 1 µL aliquot of each dilution was added to 200 µL of prepared host cells diluted in 10 mM $MgSO_4$ to an OD_{600} of 0.5. The phage and bacteria were incubated at 37°C for 15 min and added to 3 mL LB top agar (bactotryptone 10 g L^{-1}, NaCl 5 g L^{-1}, agarose 6 g L^{-1}), melted and cooled to 50°C. The LB top agar was gently mixed and plated onto LB agar plates prewarmed to 37°C. When dry, the plates were inverted and incubated at 37°C overnight. Following overnight incubation the plaques on each plate were counted and the titre in pfu ml^{-1} determined. The titres of bacteriophage stored both at 4°C and –80°C for 6 months were checked in this way.

Having calculated the titre of the amplified library it was necessary to make suitable dilutions of the bacteriophage in SM buffer, so that a suitable amount of bacteriophage could be used to infect each bacterial host strain aliquot, resulting in approximately 5 x 10^4 plaques/plate.

To plate out the library 30 large (140 mm) LB agar plates were poured, dried thoroughly in a laminar flow hood and stored at 4°C wrapped in parafilm for 48 hours. The host strain XL1-Blue MRA (P2) was prepared as detailed in Section 3.2.4.2. The bacteriophage were diluted 1 in 100 in SM buffer and 30 aliquots of 6 µL of the diluted bacteriophage were each added to each of 600 µL aliquots of prepared host cells. The bacteriophage and host cells were incubated at 37°C for 15 min and then added to 6.5 mL of LB top agar, previously melted and cooled to 50°C. The LB top agar and infected *E.coli* mixtures were then plated on to large prewarmed (37°C) LB agar plates and once dry, incubated inverted at 37°C for approximately 8 h or until the plaques reached 1-2 mm in size.

4.2.2.4 Transfer of plaques to Hybond-N membranes

Prior to transfer of plaques to nylon membranes, the plates were chilled at 4°C for a minimum of 30 min. Individual membranes were lowered onto the surface of the plates and left for 1 min while a sterile needle was used to prick perpendicularly through the membrane and plaques, marking the bottom of the Petri dish for orientation. After 1 min the membranes were carefully lifted from the LB top agar and placed plaque side up for 4 min on 2 layers of Whatmann paper soaked in denaturation solution (0.5 M NaOH, 1.5 M NaCl). The membranes were briefly blotted on dry Whatmann paper and transferred to 2 layers of Whatmann paper soaked in neutralisation solution (1.0 M Tris-HCl, pH 7.5, 1.5 M NaCl). After 4 min the membranes were again blotted briefly on dry Whatmann paper and placed on new neutralisation soaked paper for 4 min. Following this denaturation and neutralisation procedure the membranes were washed in 2 x SSC, and dried between Whatmann paper, baked at 80°C and cross-linked on an UV transilluminator (U. V. P. Inc) for 30 sec. The membranes were stored dry under vacuum until use. The original agar plates were stored wrapped at 4°C for use after screening.

4.2.2.5 Proteinase K treatment of membranes

Membranes were treated with proteinase K to degrade any cellular debris from the lysed *E. coli* cells. To achieve this, membranes were placed individually on foil, and 1.5 mL of 2 mg mL^{-1} proteinase K in 2 x SSC buffer pipetted on to the plaque side of the membrane and distributed evenly using a sterile glass spreader. The membranes were incubated at 37°C for 1 h. Following the incubation the membranes were placed between 2 sheets of Whatman paper thoroughly wetted in distilled water. The top, wetted Whatman paper sheet was gently compressed by a rolling motion with a Duran 1 L bottle. This top sheet of paper was then carefully removed taking with it any cellular or agar debris. The

process was repeated with fresh wetted Whatmann paper. The membranes were then stored either dry under vacuum or wet at 4°C until hybridisation.

4.2.2.6 Preparation of ^{32}P-labelled probes

For use in the identification of target sequences from the *C. helgolandicus* genomic library, any gene sequence must first be labelled such that it can be detected post-hybridisation. The probe can be labelled by processes such as nick translation, end-filling or random priming. The technique chosen was random priming, as this results in higher activity and therefore the detection of smaller amounts of membrane-bound DNA (Brown, 1995). The method of random primed DNA labelling originally developed by Feinberg and Vogelstein (1983 and 1984) is based on the hybridisation of random hexamers to the denatured DNA to be labelled. DNA polymerase, using the 3' OH termini of the random hexamer primers synthesizes the complimentary DNA strand incorporating labelled dNTPs present in the reaction.

DNA labelling was performed using the PRIME-IT II (Stratagene) random primed DNA labelling kit. Following gel extraction, 25 ng of the DNA to be labelled were made up to a volume of 24 µL with MilliQ water. To this DNA solution 10 µL of random oligonucleotide primers were added and the mixture heated in a boiling water bath for 5 min. Following a brief spin at room temperature, 10 µL of 5 x buffer (1 µL 0.1 mM dCTP, 1 µL 0.1 mM dGTP, 1 µL 0.1 mM dTTP), 5 µL α-^{32}P labelled dATP (3000 Ci mmol^{-1}, ICN pharmaceuticals, Inc) and 1 µL exo (-) Klenow (5 U µL^{-1}) were added, mixed briefly and incubated at 37°C - 40°C for 2-10 min. The reaction was stopped with 2 µL 0.5 M EDTA (pH 8.0). Immediately prior to use the probe was denatured by heating in a boiling water bath for 5 min.

4.2.2.7 Hybridisation of probes to Hybond-N membranes

Hybridisation solution containing 6 x SSC, 5 x Denhardts (0.1 % BSA, 0.1% Ficoll, 0.1% polyvinylpyrrolidone) and 0.5% SDS was prewarmed to 60°C. The membranes were wetted in 6 x SSC and placed in hybridisation tubes in batches of 5. To each tube 20 – 30 mL of prewarmed hybridisation solution, with the addition of denatured herring sperm DNA to a final concentration of 100 μg mL^{-1}, were added. The membranes were prehybridised at 60°C in a hybridisation oven (Techne Hybridiser HB-2D) for a minimum of one hour. The prehybridisation solution was then replaced by 10 mL of fresh hybridisation solution, prewarmed to 60°C, followed by the addition of 16 μL denatured radioactive probe. Incubation was continued overnight at 60°C with constant rotation of the hybridisation tubes.

After hybridisation the membranes were washed as follows:-

2 x 5 min washes in ~30 mL with 2 x SSC and 0.1 % SDS at room temperature

2 x 30 min washes in ~30 mL with 0.5 x SSC and 0.1 % SDS at 55°C

The membranes were sealed in Clingfilm (the number and orientation being marked with luminescent paint) and exposed to Kodak X-AR2 film backed by an intensifying screen at –80°C overnight.

4.2.2.8 Removal of probes from Hybond-N membranes

For probe removal, a solution of 0.1% (w/v) SDS was boiled, poured onto the membrane and allowed to cool to room temperature. Removal of the probe was checked by radioactive counting. Any remaining probe was stripped using an alkali procedure. This involved incubation of the membranes in prewarmed 0.2 M NaOH for 10 min at 42°C with constant agitation, followed by a second incubation with fresh alkali solution under identical conditions. The stripped membranes were rinsed in 2 x SSC for 15 min and removal of the probe checked by autoradiography.

4.2.2.9 Picking the positive plaques

The exposed films were studied for a positive hybridisation signal. Such signals on the autoradiograph were aligned (using the orientation marks) with the corresponding original agar plate. Alignment of the agar plate with the X-ray film allowed plaques within a 2.5 mm radius of the centre of the hybridisation signal to be picked with a sterile wide-bore (5 mm diameter) glass pipette. The plug of TB top agarose, containing plaques, was then expelled into a microcentrifuge tube containing 200 µL of phage buffer (20 mM Tris-HCl, pH 7.4, 100 mM NaCl, 10 mM $MgSO_4$). This process was repeated for every positive signal. The TB top agarose plugs were pooled in 200 µL of phage buffer for plates 1-15, and in a separate 200 µL of phage buffer for plates 16-30. These phage suspensions were stored at 4°C until amplified.

4.2.2.10 Amplification of the positive plaques

The picked λ-phage were amplified by re-infection into the host bacterial strain. For each probe, each library screen was amplified in two aliquots (in each case the plaques picked from plates 1-15, and plates 16-30 were amplified separately). A fresh 5 mL culture of the host strain XL1-Blue MRA (P2), in LB medium supplemented with 0.2 % maltose and 0.1 M $MgSO_4$, was grown with shaking at 37°C overnight. 500 µL aliquots of the overnight culture were then added to microcentrifuge tubes containing 40 µL aliquots of well-mixed phage suspensions. The phage and bacteria mixtures were incubated at 37°C for 20 min. 500 µL of the infected cultures were then used to inoculate 500 mL flasks containing 100 mL prewarmed (37°C) LB Broth supplemented with 0.2 % maltose and 0.1 M $MgSO_4$. The cultures were shaken at 37°C until lysis occurred (approximately 5 h). After cell lysis, 500 µL of chloroform were added and the lysate centrifuged for 10 min (RCF=8,000). The supernatant was transferred to a sterile tube and stored at 4°C.

4.2.2.11 Primary library screens

The library was first screened with the homologous developmental gene *tailless*, following the methods described in Sections 4.2.2.3 – 4.2.2.10. The membranes were then stripped with 0.1 % boiling SDS, followed by an alkali procedure (Section 4.2.2.8) and re-probed with the homologous developmental gene *even-skipped.* A fresh library was then screened with the *C. helgolandicus* homeobox clone 207 and following removal of this probe with 0.1 % boiling SDS, the library was re-probed with the *C. helgolandicus* homeobox clone 228. The process was repeated, and the membranes were probed for a third time with clone 207 at a low stringency. For the low stringency screening with clone 207 the prehybridisation and hybridisation were performed at 42°C.

4.2.2.12 Secondary and tertiary library screens

Supernatants of the amplified bacteriophage from Section 4.2.2.10 were used for a secondary screen. For each screen of the library with a different probe, a secondary screen was performed in 2 aliquots, the first including plaques picked and amplified from the original plates 1 – 15, and the second aliquot with plaques from plates 16-30. Taking into account the number of positives picked in each case and the number of plaques actually taken with each pick, the number of plaques required for a secondary screen was calculated. These calculations accounted for the under-representation of slow growing clones during the amplification, and included a 95% probability of the correct sequence being represented. With reference to previous λ amplifications the average titre of the amplified bacteriophage was assumed to be 1.6×10^9 pfu mL^{-1}. Using this assumption as a guideline, an excess of plaques was grown for a secondary screen, following the methods detailed in Section 4.2.2.3. The reduced number of plaques needed for the secondary screen enabled fewer and larger plaques to be grown per plate. The plaques were transferred to nylon membranes and then probed following the methods described in

Sections 4.2.2.4 – 4.2.2.7. The positive plaques were picked, pooled together for each of the screens with different probes, and amplified as for the primary screens (Section 4.2.2.9 and 4.2.2.10).

After the secondary screens it was still not possible to pick distinctly isolated positive clones and a tertiary screen was therefore performed. This tertiary screen allowed fewer clones to be grown to a larger size, enabling isolated positive clones to be easily picked. Six clones were picked for each of the *tailless* and Clone 207 screens and were stored individually in 100 µL phage buffer at 4°C.

4.2.3 Further analysis of the positive Lambda clones

4.2.3.1 Preparation of the recombinant Lambda DNA

The 6 positive clones picked from the tertiary screen with *tailless* and the 6 positive clones picked from the tertiary screen with clone 207 were individually amplified following the method described in Section 4.2.2.10. 10 mL of the liquid culture lysate preparations, from each of the individual amplifications, were used for the preparation of λ DNA using the Wizard® lambda preps DNA purification system (Promega UK, Ltd), following the manufacturer's instructions.

4.2.3.2 Restriction analysis of isolated clones

Lambda DNA from each isolated clone was digested with the restriction enzyme *Eco*RI to determine whether the clones selected by screening with the same probe were identical. 5 µL of prepared λ DNA for each clone were digested in a 10 µL reaction mix containing 1 µL 10 x *Eco*RI buffer H, 1 µL 1mg mL^{-1} BSA, 1µL *Eco*RI (1U μL^{-1}) and 2 µL MilliQ water. The reactions were incubated at 37°C for 1 h and 5 µL analysed by 0.6% agarose gel electrophoresis.

In order to determine which restriction fragments contained the region homologous to the gene probe Southern analysis was performed. Two digested *tailless* clones and a positive control of *Drosophila tailless* gene fragment (7-10 ng), plus two digested 207L (library) clones and a positive control of amplified clone 207 homeobox (5-10 ng), were separated by 0.8% agarose gel electrophoresis. The gel was photographed on an UV transilluminator, and the DNA transferred to Hybond-N membrane by Southern transfer (Section 2.3.18). The membrane was cut in half and the separate halves probed with the corresponding *tailless* or clone 207 α-^{32}P labelled probes, following the protocols detailed in Sections 4.2.2.6 and 4.2.2.7.

4.2.3.3 Further restriction mapping

The 207L λ clones were further digested with a range of 6-cutter restriction enzymes. These include single digests with *Eco*RI, *Sal*I, *Xba*I and *Xho*I, and double digests with *Eco*RI/*Sal*I, *Eco*RI/*Xba*I, *Eco*RI/*Xho*I, *Sal*I/*Xba*I, *Sal*I/*Xho*I, *Xba*I/*Xho*I. The digests were performed on 4 µL of prepared λ DNA from 207L clones in a total volume of 10 µL, with 5 U of enzyme(s), 1 µL of 10 x buffer D (Promega, UK Ltd) and 1 µL 1 mg mL^{-1} BSA.

The *tailless* λ clones were further digested with a range of 6-cutter restriction enzymes, all of which were compatible with buffer D (Promega UK Ltd). Digests were performed on approximately 200 ng of prepared λ DNA with 5 U of enzyme(s), 1 µL buffer D, and 1 mg mL^{-1} BSA, in a total volume of 10 µL. Single digests included *Eco*RI, *Sal*I, *Xba*I, *Xho*I, *Bam*HI, *Pst*I, and *Vsp*I. Double digests included *Eco*RI/*Sal*I, *Eco*RI/*Xho*I, *Eco*RI/*Bam*HI, *Eco*RI *Pst*I, *Sal*I/*Xho*I, *Sal*I/*Bam*HI, *Sal*I/*Pst*I, *Sal*I/*Vsp*I, *Sal*I/*Xba*I, *Xba*I/*Xho*I, *Xba*I/*Bam*HI, *Xba*I/*Pst*I.

The digests for both the *tailless* and 207L clones were separated by 1% agarose gel electrophoresis, with appropriate molecular markers. The gels were photographed on an UV transilluminator alongside a scale, and the DNA subsequently capillary blotted on to a Hybond-N membrane. The membranes were again probed with the corresponding *tailless* or clone 207 α-^{32}P labelled probes, following the protocols detailed in Sections 4.2.2.6 and 4.2.2.7. The resulting autoradiographs were compared to the original gels to determine which of the digested λ clone fragments contained the DNA sequence corresponding to the probe.

4.2.3.4 Sub-cloning and sequencing the positive λ clone fragments.

Suitable digests of both the *tailless* and 207L λ clones were repeated on a large scale. Approximately 0.5 μg of λ DNA preparation were digested with the suitable restriction enzyme(s). The digest reactions were separated by 1% agarose gel electrophoresis and the correct sized band of DNA, determined by comparison with the positive signal on the probed Southern transfer, excised. The DNA was recovered from the agarose gel using the Qiaex II gel extraction kit (Qiagen). The DNA fragments were then cloned into pBluescript SK$^-$. *E. coli* transformations were carried out as in Section 2.2.4 and recombinant clones were identified by colony PCR.

Sequencing was performed on at least three recombinant sub-clones for each DNA fragment, following the methods detailed in Sections 2.3.12-2.3.17. The insert DNA was sequenced completely in both directions. The 207L sub-clone was of a size such that it necessitated the design of two new sequencing primers, specific to the 207L sequence, in order to allow the complete sequencing of the sub-clone. The primers were designed to the central region of the 207L fragment, and allowed extended sequence data to be obtained by substituting the T3 and T7 primers traditionally used in the sequencing reactions.

Forward 207L-specific sequencing primer (T3 replacement, Tm = 62°C)

5'-gagaacaatcaagctggatgg-3'

Reverse 207L-specific sequencing primer (T7 replacement, Tm = 66°C)

5'-ccatgtcgggtccatccagc-3'

In the case in which recombinant sub-clones were found to contain one of either two similarly sized fragments, the recombinant sub-clones were first identified by colony PCR, and the correct insert subsequently identified by restricting the colony PCR products with *Sau*3AI.

4.2.4 Analysis of expression of *Calanus helgolandicus* developmental gene obtained from the library.

The temporal expression of the developmental gene obtained from the library was analysed through the construction of a solid phase cDNA library followed by PCR analysis. The method involved a modified technique described by Dynal® UK Ltd. This technique includes the construction of immobilised cDNA libraries from different *C. helgolandicus* developmental stages that can be used for multiple RT-PCR amplifications, and involves using magnetic bead separation technology. In brief, biotinylated $(dT)_{25}$ oligonucleotides (oligo-$(dT)_{25}$) are bound to streptavidin coated magnetic beads(Dynabeads® M-280 Streptavidin, Dynal Ltd). These beads can be used to obtain polyA$^+$ mRNA directly from crude lysate preparations of *C. helgolandicus* developmental stages such as eggs and nauplii. A solid phase cDNA library, specific for a particular developmental stage, can be constructed directly on the Dynabead surface using the bound oligo-$(dT)_{25}$ as a primer for the reverse transcriptase to synthesise the first strand cDNA (Figure 4.3). This results in a covalently linked first strand cDNA library that can be used for cDNA amplification.

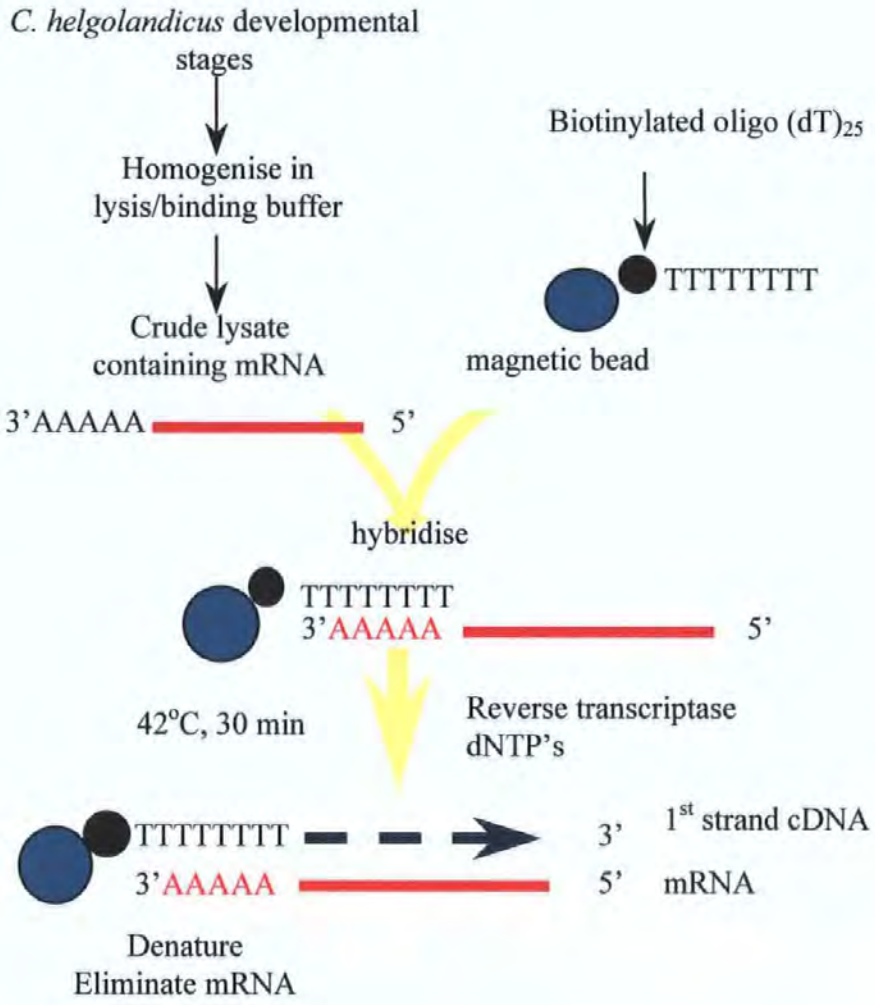


Figure 4.3 Diagrammatic representation of the construction of a solid phase cDNA library from different *Calanus helgolandicus* developmental stages.

The technique has many advantages including the fact that only a small amount of cDNA Dynabeads is necessary for PCR amplification. The solid-phase cDNA library is also re-usuable, allowing multiple copies of the cDNA of a specific mRNA (second-strand cDNA) to be generated using a single-sided PCR with a specific primer. The second strand is melted off from the solid phase template, the Dynabeads recovered by magnetic separation, and the supernatant with the second strand cDNA used for amplification (Figure 4.4).

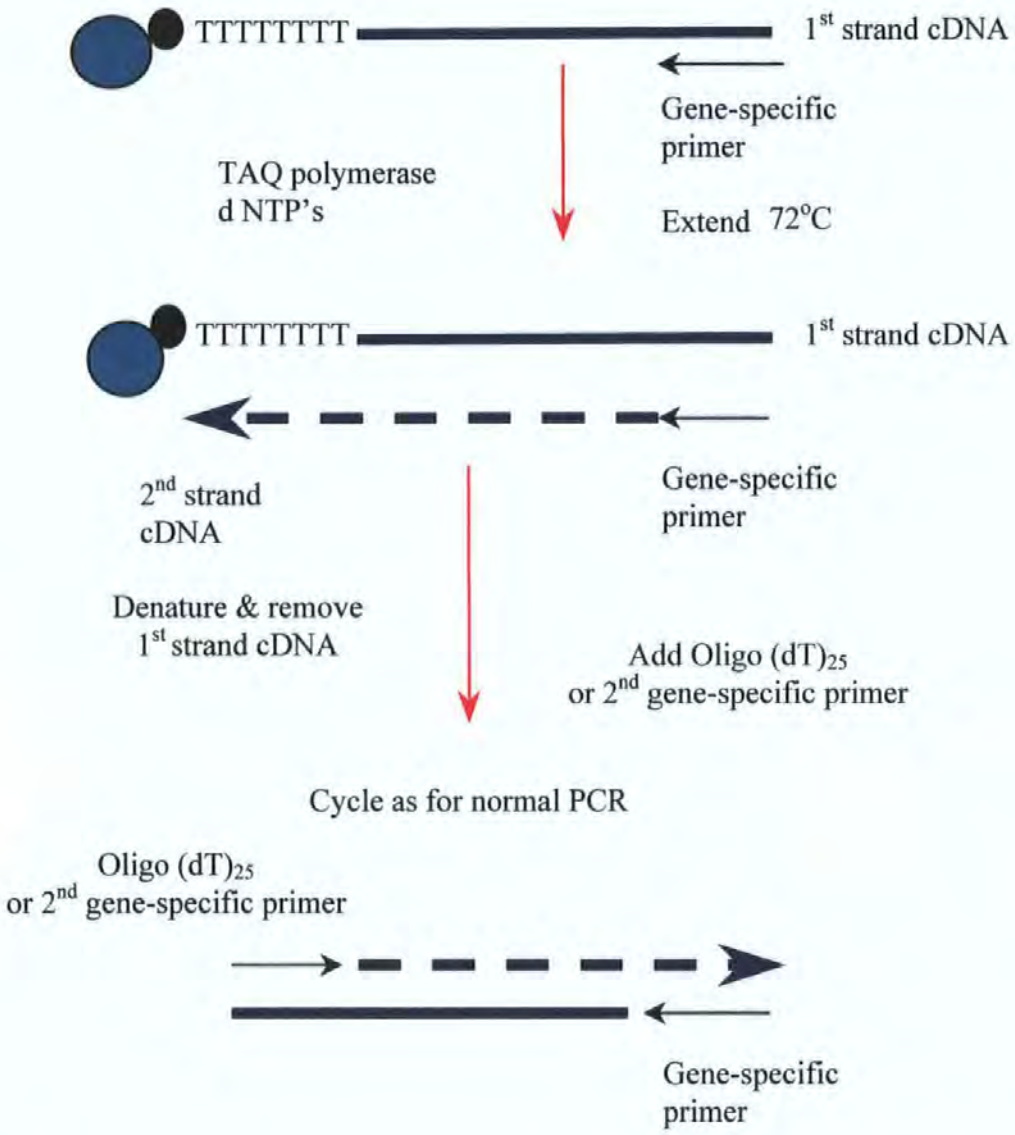


Figure 4.4 Diagrammatic representation of PCR from solid phase cDNA library.

4.2.4.1 Preparation of oligonucleotide $(dT)_{25}$ Dynabeads

Biotinylated $(dT)_{25}$ oligonucleotides (oligo-$(dT)_{25}$) were bound to streptavidin coated magnetic beads(Dynabeads® M-280 Streptavidin, Flowgen Instruments Ltd) for the direct isolation of mRNA from crude lysate. The Dynabeads were first prepared for RNA manipulation by washing, with the aid of a magnetic separation stand, twice with 2

volumes solution A (DEPC treated 0.1 M NaOH, 0.05 M NaCl) and once with 2 volumes solution B (DEPC treated 0.1 M NaCl). The beads were then re-suspended in 2 volumes solution B, giving a final concentration of 5 μg μL^{-1}.

For each isolation approximately 325 μg of Dynabeads were used. This quantity of Dynabeads has the capacity to bind approximately 66 pmoles of a single stranded oligonucleotide. Therefore 3.5 μL of biotinylated $(dT)_{25}$ (150 ng μL^{-1}) were bound to 325 μg of streptavidin coated Dynabeads as follows: The Dynabeads were washed once in 66 μL 2 x B+W buffer (DEPC treated 2.0 M NaCl, 1 mM EDTA, 10 mM Tris-HCl, pH 7.5) and then re-suspended in the same volume of 2 x B+W buffer. 66 pmoles (3.5 μL) of biotinylated oligo-$(dT)_{25}$ were made to an equal volume with DEPC treated MilliQ water and added to the washed streptavidin coated Dynabeads. The oligonucleotide and beads were incubated at room temperature for 10-15 min with gentle mixing. Following the incubation the beads and adhered biotinylated oligo-$(dT)_{25}$ were washed three times with 2 x B+W buffer.

4.2.4.2 Preparation of Calanus helgolandicus *lysates from specific developmental stages*

Eggs were collected from a culture of *C. helgolandicus* by removal of female *C. helgolandicus*, contained in a plexiglass tube with 300 μm mesh bottom, from a 5-L beaker of filtered sea water. The remaining eggs were filtered through a 53 μm mesh and then washed into a Petri dish. The Petri dish was gently swirled to aggregate the eggs in the centre of the dish. Without disturbing the eggs, as many faecal pellets and residual algae as possible were removed with a pipette under a dissecting microscope. The eggs were then transferred to a clean Petri dish by pipette. This process was repeated several times until the eggs were isolated from any contaminating material, with residual faecal pellets and food being removed with a dissecting needle.

Over 1800 eggs were obtained from the culture in this way. The eggs were then divided into three microcentrifuge tubes with 300 eggs (approximately 10 mg), 500 eggs (approximately 6.5 mg) and 1000 eggs (approximately 10 mg) being placed into individual tubes. These were centrifuged for 2 min (RCF=11600) and any excess seawater removed. To each tube, 250 µL of lysis/binding buffer (DEPC treated 100 mM Tris-HCl, pH 8.0, 50 mM LiCl, 10 mM EDTA, pH 8.0, 1% LiDs, 5 mM dithiotreitol) were added. The *C. helgolandicus* eggs were then homogenised using a pellet pestle homogeniser (Anachem Ltd.). The homogenates were centrifuged for 1 min (RCF=11600) and the supernatant transferred to a clean tube. The supernatants were stored either on ice for up to 30 min, or frozen in liquid nitrogen and stored at –80°C.

Further lysates were obtained from specific developmental stages. Eggs > 18h old and eggs < 18h old were obtained by removing the adult female *C. helgolandicus* in their plexiglass tube from one 5 L beaker of filtered seawater to a 5 L beaker of fresh filtered seawater. The culture was continued at 15°C and the original beaker stored overnight at the same temperature. The following morning eggs from both the old beaker (>18 h old) and from the new beaker (<18 h old) were collected as before. Approximately 500 eggs over 18 h old, and 500 eggs less than 18 h old (equating to approximately 6.5 mg of tissue each) were isolated in this way and transferred to a microcentrifuge tube.

C. helgolandicus nauplii, copepodites and adults were also picked from culture. Approximately 300 nauplii including stages NI-NVI (approximately 10 mg), 8 varying copepodite stages, from CI - CV (approximately 10 mg) and 5 adults, mixed male and female (approximately 10mg) were placed in RNase free microcentrifuge tubes and the crude lysate prepared as above.

4.2.4.3 Direct mRNA isolation from crude lysate

The mRNA was isolated from the prepared crude lysates with the oligonucleotide $(dT)_{25}$ Dynabeads. 325 µg of oligo $(dT)_{25}$ Dynabeads were used to isolate between 400 – 800 ng of poly A^+ mRNA from each crude lysate preparation.

Before use, the pre-washed oligo $(dT)_{25}$ Dynabeads were washed once in 100 µL lysis/binding buffer. The lysis/binding buffer was removed and each crude lysate preparation added to an aliquot of oligo $(dT)_{25}$ Dynabeads. The poly A^+ tail was allowed to anneal to the oligo $(dT)_{25}$ by incubation, with constant rotation at room temperature for 5 min. The supernatant was removed and the beads washed twice with 200 µL washing buffer + LiDS (DEPC treated 10 mM Tris-HCl, pH 8.0, 0.15 M LiCl, 1 mM EDTA, 0.1% LiDS) and once with 200 µL washing buffer (DEPC treated 10 mM Tris-HCl, pH 8.0, 0.15 M LiCl, 1 mM EDTA). The beads were finally washed three times with 250 µL RT-buffer (10 mM Tris-HCl, pH 8.3, 75 mM KCl) being transferred to a new tube between washing steps 2 and 3.

4.2.4.4 Synthesis of solid phase cDNA library

After the third wash, the RT buffer was removed and the beads re-suspended in 50 µl of reverse transcription reaction mix [10 µL of 5 x reaction buffer, 10 µL 10 mM dNTPs, 0.5 µL rRNasin (20-25 U μL^{-1}), 1 µL Moloney Murine Leukaemia Virus reverse transcriptase (M-MLV RT; 200 U μL^{-1}) and 28.5 µL DEPC treated MilliQ water]. The beads were mixed gently and incubated at room temperature for 10 min, followed by an incubation of 1 h at 42°C to permit synthesis of the first strand cDNA. The beads were then collected, and the reverse transcription mix removed and replaced by 50 µL TE buffer. This was heated to 95°C for 1 min, allowing the beads and annealed cDNA to be collected magnetically and the supernatant containing the melted mRNA to be removed

and discarded. The beads were washed once in TB-buffer (20 mM Tris-HCl, pH 8.4, 50 mM KCl, 1.5 mM $MgCl_2$, 0.1 mg mL^{-1} BSA) and then stored in 50 µL TB-buffer at 4°C. The single-stranded cDNA library was found to be stable in this form (in nuclease-free solutions) for at least 2 months.

4.2.4.5 PCR analysis of the solid phase cDNA library with one gene-specific primer and an oligo $(dT)_{25}$

Temporal expression of the *C. helgolandicus* developmental gene obtained from the genomic library (207L) was analysed by amplification of this gene from the solid phase cDNA library, prepared from *C. helgolandicus* eggs. A gene specific forward primer was designed to the 207L fragment, for the initial amplification of the second strand cDNA.

207L-RTIb 5'-aggaagcgaggtcgcc-3'

207L-RT1b was designed to the start of the homeobox sequence. The forward primer 207L-RT1b was used with a reverse primer (oligo $(dT)_{25}$), complimentary to the poly A^+ tail of the first strand cDNA.

For the reaction a 50 µg aliquot of the solid phase cDNA library was prepared by washing in 1 x *Taq* DNA polymerase buffer (10 µL) with the use of a magnetic separation stand (Promega UK Ltd). The *Taq* DNA polymerase buffer was then removed and the beads re-suspended in 50 µL of PCR mix:-

Dynabeads with first strand cDNA	- (50 µg)
100 ng μL^{-1} forward primer (207L-RTI or 207L-RTIa)	- 5 µL
100 ng μL^{-1} oligo $(dT)_{25}$ primer	- 5 µL
2 mM dNTP's	- 5 µL
10 x *Taq* DNA polymerase buffer	-5 µL
Taq DNA polymerase	- 0.5 µL
MilliQ water	- 29.5 µL

A further 50 µL of the PCR mix were used as a negative control, omitting any template DNA. Amplifications were carried out in a thermal cycler (PTC-100™, MJ Research, Inc.). The cycling parameters involved an initial cycle including a denaturation step of 94°C (1 min), an annealing step of Tm (1 min) and an extension step of 72°C (5 min) to ensure the complete synthesis of second strand cDNA. This initial cycle was followed by a denaturation step at 94°C (4 min) which was paused briefly to permit removal of the Dynabeads with bound first strand cDNA. The removal was achieved by retention of the beads with a magnetic separation stand and rapid transfer of the remaining PCR mix, containing the second strand cDNA to a clean PCR tube. The amplification was then continued with 35 cycles of Tm (1 min), 72°C (1.5 min) and 94°C (1 min). A final annealing phase at Tm (2 min) was followed by an extension phase at 72°C (5 min) and storage at 4°C until use.

The Tm used in the amplification cycle was varied from 40°C to 57°C in order to obtain an optimum PCR band. The amplification products were analysed by agarose gel electrophoresis and viewed on an UV transilluminater.

Potentially positive bands were capillary blotted on to a Hybond –N membrane by Southern transfer (Section 2.3.18). The membrane was then probed with both the original clone 207 homeobox fragment, used to probe the library, and the 207L sub-clone sequence. The sequences were labelled with α-^{32}P (Section 4.2.2.6) and used to probe the membrane following the hybridisation protocol in Section 4.2.2.7. The amplification product found to give a positive hybridisation signal when probed was further analysed by cloning into pBluescript SK$^-$ and sequencing (Section 2.3).

4.2.4.6 PCR analysis of the solid phase cDNA libraries with two gene-specific primers

It proved necessary to use 2 gene-specific primers to analyse the temporal expression of 207L. Amplification from cDNA with 2 gene-specific primers is well established, for example Damen *et al.* (1998). The forward primer 207L-RTIa, and a reverse primer designed to the end of the 207L sub-clone (207L-RTII), were used.

207L-RTII 5'-gactctggtgactcgtcc-3'

These 2 primers were used to amplify 50 µg cDNA library constructed from *C. helgolandicus* egg lysates. The aliquot of cDNA library was washed in 1 x *Taq* DNA polymerase buffer before re-suspending in the following PCR reactant mix:-

Dynabeads with first strand cDNA	- (50 µg)
100 ng μL^{-1} forward primer, 207L-RTIa	- 5 µL
2 mM dNTP's	- 5 µL
10 x *Taq* DNA polymerase buffer	-5 µL
MilliQ water	- 25 µL

The cycling parameters were altered from those detailed in Section 4.2.4.5 by including a hot start (delayed addition of *Taq* DNA polymerase), and delayed addition of the reverse primer. The initial cycle including a denaturation step of 94°C (1 min), an annealing step of 62°C (1 min) and an extension step of 72°C (5 min), was paused at the end of the denaturation step for the addition of 0.5 µl *Taq* DNA polymerase diluted in 4.5 µL MilliQ water. This initial cycle was followed by a denaturation step at 94°C (4 min) which was briefly paused to allow removal of the Dynabeads and bound first strand cDNA as before. At this point 5 µL of the reverse primer 207L-RTII were added to the reaction mixture. The amplification was then continued with 35 cycles of 62°C (1 min), 72°C (1 min) and 94°C (1 min). A final annealing phase at 62°C (2 min) was followed by an extension phase at 72°C (5 min) and storage at 4°C until use.

A negative control with no DNA template was included and the resulting amplification products were analysed by 1 % agarose gel electrophoresis and viewed on an UV transilluminator. The amplifications were repeated on the cDNA libraries obtained from other *C. helgolandicus* developmental stages, following exactly the same protocol as above.

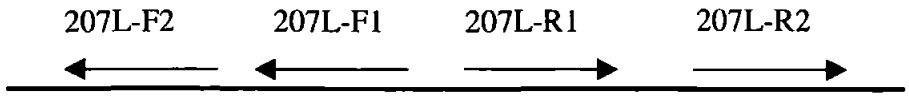
The identity of the distinct bright DNA band, obtained with the two gene-specific primers, was confirmed by cloning into pBluescript SK$^-$ and sequencing (Section 2.3).

4.2.5 Obtaining sequence data outside the 207L sub-clone by Inverse PCR

To obtain information outside the 207L sub-clone a technique known as inverse PCR (IVPCR) was applied to the original 207L λ clone. This advanced PCR method has been developed to allow the amplification of sequences that lie outside the boundaries of known sequences (Ochman *et al.*, 1988; Triglia *et al.*, 1988). Further details of IVPCR are given in Chapter 5, but in brief, the technique involves restricting the λ clone then re-ligating at a low concentration to promote the formation of circular molecules. Two pairs of divergent primers were designed, specific to the 207L sub clone, the 207L sub clone representing a small, previously sequenced region of the λ clone. The primer pairs were.

207L-F1	5'-cgaaaacgacagcgaacc-3'
207L-R1	5'-ccgagtgtaggtttggcg-3'
207L-F2	5'-actgacgaaccagaccag-3'
207L-R2	5'-acctcgcttcctctctg-3'

A schematic representation of the relative positions of the 207L IVPCR primers on the 207L sub-clone is shown below.



500 ng aliquots of the 207L λ DNA preparation were digested with the endonuclease restriction enzymes *Xba*I, *Eco*RI, *Dde*I, *Vsp*I, *Hind*III, *Eco*RV, and *Bam*HI (12.5 μL λ DNA preparation, 1.5 μL appropriate 10 x buffer, 1 μL enzyme), at 37°C for 3 h. The digest reactions were then made to a volume of 400 μL with MilliQ water and extracted once with phenol/chloroform:isoamyl alcohol 24:1, once with chloroform:isoamyl alcohol 24:1, and then precipitated with 1/10 volume 3M NaOAc and 2.5 volume 100% ethanol at –80°C for 1 h. The washed and dried pellets were re-suspended in 132.4 μL MilliQ water, 15 μL 10 x ligase buffer and 2.6 μL T_4 DNA ligase (2 U μL^{-1} Promega UK Ltd.). The ligations were then incubated at 16°C overnight. Following heat inactivation at 70°C for 10 min, the circularised λ DNA was stored at 4°C for use in the IVPCR reactions.

For each of the 7 restricted and circularised stock solutions 2 μL were used as the template DNA in an IVPCR reaction containing the following reactants:-

Aliquot of circularised DNA	- 2 μL
(100 ng μL^{-1}) Primer 207L-F1	- 2.5 μL
(100 ng μL^{-1}) Primer 207L- R1	- 2.5 μL
10 x *Taq* DNA polymerase buffer	- 2.5 μL
25 mM MgCl	- 1.0 μL
2 mM dNTP's	- 2.5 μL
Taq DNA polymerase	- 0.5 μL
Single stranded binding protein (2.2 mg mL^{-1})	- 0.08 μL
MilliQ water	- 11.42 μL

Amplification was carried out in a thermal cycler (PTC-100™, MJ Research, Inc.). The cycling parameters involved an initial denaturation step of 94°C (4 min), followed by 35 cycles of 55°C (30 sec), 72°C (1 min) and 94°C (30 sec). A final annealing step of 55°C (1 min) was followed by a long extension step at 72°C (5 min) and storage at 4°C.

A 1 μL aliquot of the PCR product was used directly in a second, nested IVPCR, using 207L-F2 and 207L-R2 primers. The reaction mix was as follows:-

Initial IVPCR product	- 1.0 μL
(100 ng μL^{-1}) Primer 207L-F2	- 2.5 μL
(100 ng μL^{-1}) Primer 207L-R2	- 2.5 μL
10 x *Taq* DNA polymerase buffer	- 2.5 μL
25 mM MgCl	- 1.0 μL
2 mM dNTP's	- 2.5 μL
Taq DNA polymerase	- 0.5 μL
MilliQ water	- 12.5 μL

Amplification was performed using the same parameters as for the initial IVPCR. Negative controls were included for both the initial and nested IVPCR. 10 μL of each PCR product were analysed by 0.8 % agarose gel electrophoresis. The positive IVPCR amplification was cloned into pBluescript SK$^-$, following the methods in Section 2.3.8-2.3.11. The sequence was determined commercially by MWG biotech (UK) limited.

The sequence data was analysed to locate the position of the original restriction site and then rearranged to give the contiguous genome sequence. The central region of the 207L sub-clone was incorporated to complete the sequence between the IVPCR primers.

4.3 Results

4.3.1 Probes prepared to screen the library

The following homologous developmental genes, in preparation for the construction of probes, were successfully amplified from *Drosophila melanogaster* genomic DNA using primers designed from sequences on the computer database.

Gene	Size of amplified product	Molecular nature
tailless	826 bp	Zinc finger
giant	826 bp	Leucine zipper
kruppel	229 bp	Zinc finger
orthodenticle	1025 bp	Homeobox
hairy	784 bp	Helix-loop-helix
fushi tarazu	525 bp	Homeobox
runt	655 bp	Nuclear protein
even skipped	804 bp	homeobox

Table 4.1 Homologous developmental gene fragments amplified from *Drosophila* melanogaster genomic DNA in preparation for the construction of probes.

4.3.2 Library Screening

4.3.2.1 Titre of stored library

Prior to plating out the library for screening, the titres of the amplified and stored bacteriophage were established. The average titre of the serially diluted bacteriophage stored at –80°C, with the addition of 7 % v/v DMSO, was 8.13 x 10^8 pfu mL^{-1}. The average titre of the serially diluted bacteriophage stored at 4°C, with 0.3 % v/v chloroform, was 8.20 x 10^8 pfu mL^{-1}. There was no significant difference in titre between either of the stored libraries.

With an average titre of 8.2 x 10^8 pfu mL^{-1}, it was calculated that 0.06 µL of bacteriophage were required to infect 600 µL aliquots of prepared host cells in order to produce 5 x 10^4 plaques/plate. 30 aliquots of 6 µL diluted bacteriophage (1 in 100 with SM buffer) were therefore used to infect 30 aliquots of 600 µL host bacterial cells, to create a library of 1.5 x 10^6 plaques.

4.3.2.2 Primary library screens

The number of potentially positive hybridisation signals, observed and picked, for each screen of the library with a different probe are shown in Table 4.2. Taking into account that for each positive signal the number of plaques actually picked from the plate with a 2 mm bore pipette is approximately 50, the number of plaques needed after amplification for a successful secondary screening is given. This calculation accounts for the potential of under-representation of slow growing clones during amplification and gives a 95% probability of each clone being represented.

Probe		Number of positive hybridisation signals	Total number of plaques picked	Number of plaques needed for secondary screen
tailless	Plates 1-15	25	1250	4000
	Plates 15-30	21	1050	3200
even skipped	Plates 1-15	4	200	600
	Plates 15-30	3	150	450
clone 207	Plates 1-15	29	1450	4500
(↑ stringency)	Plates 15-30	22	1100	3500
clone 228	Plates 1-15	2	100	300
	Plates 15-30	1	50	150
clone 207	Plates 1-15	-	-	-
(↓ stringency)	Plates 15-30	-	-	-

Table 4.2 Results of primary library screens and calculation of the number of plaques required for a secondary screen.

4.3.2.3 Secondary and tertiary library screens

Table 4.3 shows the number of putative positive hybridisation signals seen when a secondary screen was performed with *tailless*, *even-skipped*, clone 207 and clone 228.

Taking into account amplification of the plaques and given a 95% probability of each clone being represented, the number of plaques needed for a tertiary screen is shown.

Probe	Number of positive hybridisation signals	Total number of plaques picked	Number of plaques needed for tertiary screen
tailless	11	18	54
even skipped	-	-	-
clone 207 (↑ stringency)	21	252	800
clone 228	-	-	-

Table 4.3 Products of secondary library screens, and calculation of plaque numbers required for a tertiary screen.

The secondary screens with both the *tailless* and clone 207 probes resulted in definite, clear hybridisation signals, meriting a tertiary screen to obtain isolated clones. The *even-skipped* and clone 228 secondary screens produced no clear hybridisation signals and were not analysed further.

The tertiary screens with *tailless* and clone 207 produced many clear and distinct hybridisation signals. These positive signals were easily aligned with distinct isolated plaques on the corresponding plates, allowing the acquisition of 6 isolated λ clones from both the *tailless* and clone 207 screenings.

Single representative autoradiographs for both the *tailless* and clone 207 primary, secondary and tertiary screens are shown in Figure 4.5. This Figure illustrates the progressive size and number of positive hybridisation signals obtained during the three screening processes.

Hox 207 screenings

1^0 Screening	2^0 Screening	3^0 Screening
30 plates	3 plates	4 plates
~50,000 plaques / plate	~7,000 plaques / plate	~300 plaques / plate

Tailless screenings

1^0 Screening	2^0 Screening	3^0 Screening
30 plates	3 plates	2 plates
50,000 plaques / plate	~1,600 plaques / plate	~ 600 plaques / plate

Figure 4.5 A representative autoradiograph for each of the three screens with the *tailless* and clone 207 probes. Large signals at the circumference are orientation marks. Small ticks seen on the primary screens indicate that the plaque was picked

The difference in intensity of the hybridisation signal between the clone 207 homologous probe and the *tailless* heterologous probe can clearly be seen in the tertiary screens, with the clone 207 homologous probe signal being more intense than that of the *tailless* heterologus probe.

4.3.3 Further analysis of the positive *tailless* and 207L λ clones

4.3.3.1 Analysis of the λ clones by restriction digest with Eco*RI and Southern analysis*

The 6 λ clones obtained from the library screen with *tailless* proved to be similar, with each clone resulting in matching restriction patterns when digested with *Eco*RI. The 6 λ clones obtained from the library screen with the clone 207 homeobox were also proven to be similar, with identical restriction profiles being obtained when digested to completion with *Eco*RI.

To determine the region of the λ clone inserts corresponding to the probes, 2 *Eco*RI digested *tailless* λ clones and 2 *Eco*RI digested 207L λ clones were capillary blotted onto Hybond-N membrane and probed with the corresponding radiolabelled *tailless* and clone 207 probes. Figure 4.6 shows the agarose gel of the digested λ clones, including a positive control for both *tailless* and clone 207.

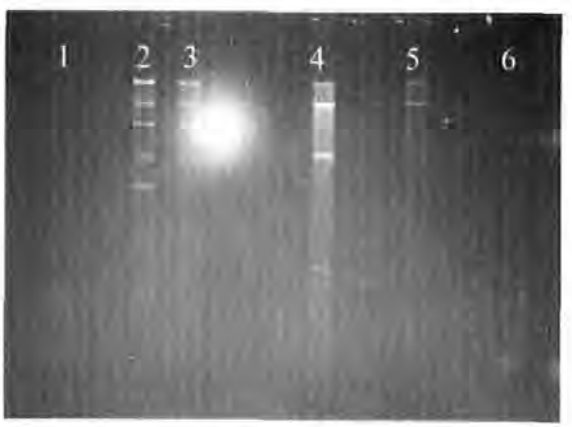


Figure 4.6 0.6% agarose gel showing the restriction profiles of *Eco*RI digested *tailless* and 207L λ clones. Lane 1 = 7-10 ng *tailless* positive control (*Drosophila melanogatser tailless* gene fragment); lanes 2 and 3 = 5 μL *tailless* λ clones digested with *Eco*RI; lanes 4 and 5 = 5 μL and 2 μL respectively of 207L λ clones digested with *Eco*RI; lane 6 = 5-10 ng clone 207 positive control (amplified clone 207 homeobox fragment).

The results of the Southern analysis are shown in Figure 4.7.

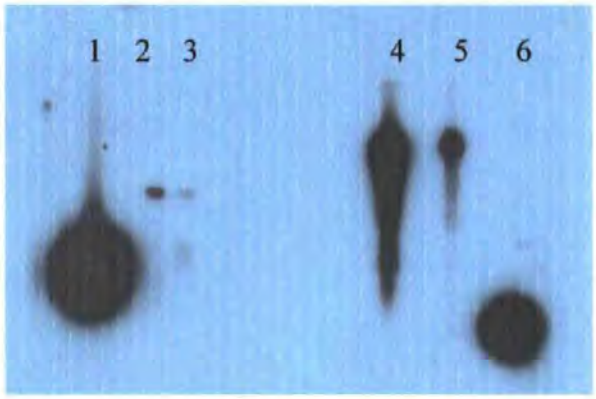


Figure 4.7 Southern analysis of *Eco*RI digested *tailless* and 207L λ clones probed with the corresponding radiolabelled *tailless* and clone 207 probes. Lane 1 = 7-10 ng *tailless* positive control (*Drosophila melanogatser tailless* gene fragment); lanes 2 and 3 = *tailless* λ clones digested with *Eco*RI; lanes 4 and 5 = 207L λ clones digested with *Eco*RI; lane 6 = 5-10 ng clone 207 positive control (amplified clone 207 homeobox fragment).

From the autoradiograph it is clear that both the *tailless* and clone 207 positive controls have produced a strong hybridisation signal. The *tailless* fragment which hybridised to the *tailless* probe, as shown on the autoradiograph, corresponds to the 2.3 kb band of the *Eco*RI digested *tailless* λ clone. The autoradiograph also clearly shows that the positive hybridisation signal to the clone 207 probe corresponds to the 9 kb band of the *Eco*RI digested 207L λ clone.

The band of the *Eco*RI digested 207L λ clone, which by Southern analysis was found to hybridise to the clone 207 probe, was of a similar size to the smaller 9 kb arm of the vector. As these two fragments co-migrate on an agarose gel it was impossible to sub-clone the correct fragment. Therefore it was necessary to further restriction map and analyse the 207L λ clone in order to obtain a suitable fragment to sub-clone.

The restricted *tailless* λ clone fragment found to correspond to the *tailless* probe, although of a suitable size, proved to be unsuitable for sub-cloning. Any attempts to clone either part of, or the complete 2.3 kb fragment, proved the sequence resistant to sub-cloning. It could only be assumed that the fragment contained a "poison" sequence preventing cloning, and therefore further analysis of the *tailless* λ clone was also needed to obtain a fragment suitable for sub-cloning.

4.3.3.2 Further restriction mapping and Southern analysis

To find fragments of the *tailless* and 207L λ clones of a suitable size for sub-cloning, and which correspond to the appropriate probes, the λ clones were digested to completion with a range of restriction endonucleases. The restriction digests were fractionated by 1% agarose gel electrophoresis, capillary blotted onto Hybond-N membranes and the Southern transfers analysed.

Figures 4.8 and 4.9 show the fractionation of the restriction digests for the *tailless* and 207L λ clones respectively. The results of the Southern transfer of the *tailless* λ clone digests, following probing with the *Drosophlia melanogaster tailless* gene fragment probe (Section 4.2.1.1), are shown in Figure 4.10. The results of the Southern transfer of the 207L λ clone digests, following probing with the *C. helgolandicus* clone 207 homeobox fragment (Section 4.2.1.2) are shown in Figure 4.11.

Figure 4.8 Fractionation of the *tailless* λ clone, digested with a range of restriction endonucleases, by 1% agarose gel electrophoresis. Lane 1 = 200 ng kb ladder; lane 2 = 200 ng bp ladder; lane 3 = 200 ng λ *Hin*dIII; lane 4 = 200 ng Phi X174/*Hae*III. Lanes 5-23 = *tailless* λ clones digested with the following restriction enzymes:- lane 5 = *Eco*RI; lane 6 = *Sal*I; lane 7 = *Xba*I; lane 8 = *Xho*I; lane 9 = *Bam*HI; lane 10 = *Pst*I; lane 11 = *Vsp*I; lane 12 = *Eco*RI/*Sal*I; lane 13 = *Eco*RI/*Xho*I; lane 14 = *Eco*RI/*Bam*HI; lane 15 = *Eco*RI/*Pst*I; lane 16 = *Sal*I/*Xho*I; lane 17 = *Sal*I/*Bam*HI; lane 18 = *Sal*I/*Pst*I; lane 19 = *Sal*I/*Vsp*I; lane 20 = *Sal*I/*Xba*I; lane 21 = *Xba*I/*Xho*I; lane 22 = *Xba*I/*Bam*HI; lane 23 = *Xba*I/*Pst*I; lane 24 = 200 ng Phi X174/*Hae*III, lane 25 = 200 ng λ *Hin*dIII.

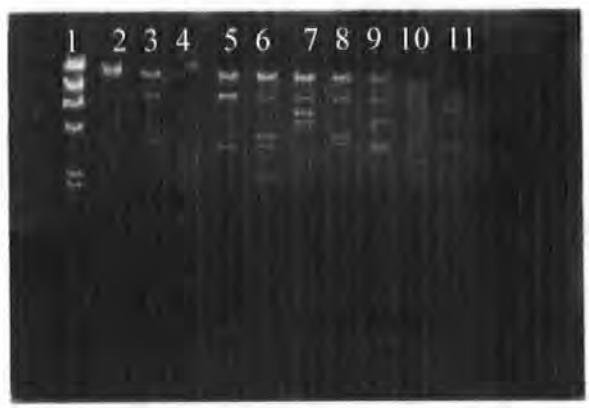


Figure 4.9 1% agarose gel electrophoresis of 207L λ clones digested with a range of restriction endonucleases:- Lane 1 = 200 ng λ *Hin*dIII. Lanes 2-11 = 207L λ clones digested with the following restriction enzymes:- lane 2 = *Xho*I; lane 3 = *Xba*I; lane 4 = *Sal*I; lane 5 = *Eco*RI; lane 6 = *Xba*I/*Xho*I; lane 7 = *Sal*I/*Xho*I; lane 8 = *Sal*I/*Xba*I; lane 9 = *Eco*RI/*Xho*I; lane 10 = *Eco*RI/*Xba*I; lane 11 = *Eco*RI/*Sal*I.

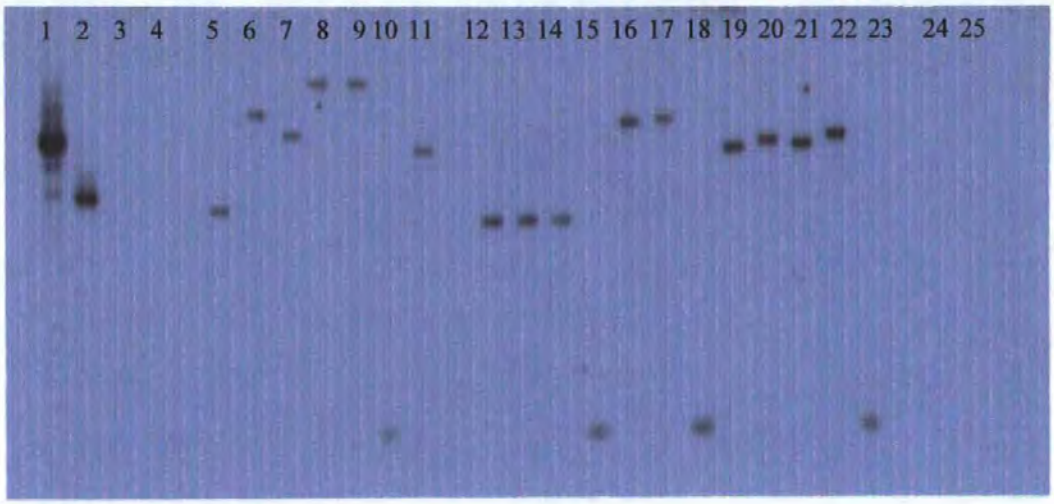


Figure 4.10 Southern analysis of *tailless* λ clone digests, probed with *tailless* probe. Lane 1 = 200 ng kb ladder; lane 2 = 200 ng bp ladder; lane 3 = 200 ng λ *Hin*dIII; lane 4 = 200 ng Phi X174/*Hae*III. Lanes 5-23 = *tailless* λ clones digested with the following restriction enzymes:- lane 5 = *Eco*RI; lane 6 = *Sal*I; lane 7 = *Xba*I; lane 8 = *Xho*I; lane 9 = *Bam*HI; lane 10 = *Pst*I; lane 11 = *Vsp*I; lane 12 = *Eco*RI/*Sal*I; lane 13 = *Eco*RI/*Xho*I; lane 14 = *Eco*RI/*Bam*HI; lane 15 = *Eco*RI/*Pst*I; lane 16 = *Sal*I/*Xho*I; lane 17 = *Sal*I/*Bam*HI; lane 18 = *Sal*I/*Pst*I; lane 19 = *Sal*I/*Vsp*I; lane 20 = *Sal*I/*Xba*I; lane 21 = *Xba*I/*Xho*I; lane 22 = *Xba*I/*Bam*HI; lane 23 = *Xba*I/*Pst*I; lane 24 = 200 ng Phi X174/*Hae*III, lane 25 = 200 ng λ *Hind*III.

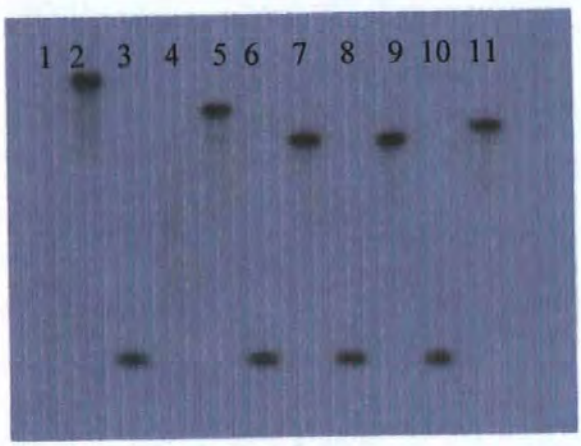


Figure 4.11 Southern analysis of 207L λ clone digests. Lane 1 = 200 ng λ *Hind*III. Lanes 2-11 = 207L λ clones digested with the following restriction enzymes:- lane 2 = *Xho*I; lane 3 = *Xba*I; lane 4 = *Sal*I; lane 5 = *Eco*RI; lane 6 = *Xba*I/*Xho*I; lane 7 = *Sal*I/*Xho*I; lane 8 = *Sal*I/*Xba*I; lane 9 = *Eco*RI/*Xho*I; lane 10 = *Eco*RI/*Xba*I; lane 11 = *Eco*RI/*Sal*I.

From the autoradiograph (Figure 4.10) of the *tailless* restriction digests it is clear that the fragment of DNA that hybridised to the *tailless* probe, which proved to be the most suitable size for further sub-cloning, corresponds to the approximately 300 bp band of the *tailless* λ clone *Pst*I digest (lanes 10, 15, 18, 23). The next fragment of DNA in size that also hybridised to the *tailless* probe corresponds to the approximately 2.3 kb band of the *tailless* λ clone *Eco*RI digest (lane5, lanes 12-14) that had previously proved resistant to further sub-cloning. The next size fragment again corresponds to the approximately 6 kb band of the *tailless* λ clone *Xba*I digest (lanes 7, 20-22). Both the approximately 300 bp band of the *Pst*I digest and the approximately 6 kb band of the *Xba*I digest were sub-cloned.

The autoradiograph of the 207L digested λ clones (Figure 4.11) clearly shows that the most appropriately sized band of DNA that hybridised with the clone 207 homeobox probe is the approximately 400 bp band of the 207L λ clone *Xba*I digest. This band was further analysed by sub-cloning.

4.3.3.3 Sub-cloning and sequencing

The large *Xba*I *tailless* (approximately 6 kb) fragment again proved resistant to cloning. The approximately 300 bp DNA fragment of the *tailless* λ clone *Pst*I digest was successfully cloned. However, sequencing showed that in fact 2 similarly sized fragments of different origin had been cloned. *Pst*I is not unique to the polylinker of the vector and therefore, fragments of the λ DASHII vector were also combined with the restriction fragments of the insert DNA. Unfortunately one such vector fragment co-migrated with the approximately 300 bp insert fragment that hybridised to the *tailless* probe. The clones however, were identified to be of vector or insert origin by digesting recombinant plasmid preparations with *Sau*3AI. A distinct restriction pattern was seen for the clones of vector

origin and a distinct restriction pattern was seen for the clones containing the fragment of insert DNA that hybridised to the *tailless* probe. Three clones identified as containing the correct fragment were then sequenced.

The sequence of the sub-clone that hybridised to the *tailless* probe is shown in Figure 4.12. The 'bestfit' (region of highest homology between 2 sequences) of the *tailless* sub-clone with the *Drosophila tailless* gene is shown in Figure 4.13.

```
1     atgggggcga gctctcctac caactctttc ccaagatagt tgcaacattc
51    ctgggcttcc aatctcatgt tgccccaggc ttggaggaac tcttcagcgg
101   tccggctgcc accttccagg aaagttcccc acctgttctg acccccctcga
151   cactgcccac cacctcctcc agctgatggc agatgccatg ctcccctgtg
201   aagtggggaa gagacatctc aatgcccccc atgaatgcag ctgcactggt
151   ttcagcctgg gagcggagac ccagcccacc cagcttgaca ggctgatgga
201   ccagccactg ctggaagctc ctgccc
```

Figure 4.12 Sequence of the *Calanus helgolandicus* sub-clone that hybridised to the *tailless* probe.

```
Calanus tailless      134   tgttctgacccccctcgacactgcccaccacctcct
                            ||  ||| || |   | | |||||| ||| |||||
Drosophila tailless   565   tgccctggccactcgggccctgcccccca.ctcct

Calanus tailless      169   ccagctgatggcagatg
                            || ||||||||||| ||
Drosophila tailless   600   cc.gctgatggcagctg
```

Figure 4.13 'Bestfit' of the *tailless* subclone from *Calanus helgolandicus* with the *Drosophila* melanogaster *tailless* gene. Vertical lines show the homology between the 2 sequences, . = alignment gap

The sequence data obtained from the 226 bp *C. helgolandicus tailless* subclone shows a 'bestfit' with high homology to the *Drosophila* sequence. This homology is

confined to a small region however, and comparison of the complete sub-clone indicates a low level of homology with the *Drosophila melanogaster tailless* sequence, both at the nucleotide and amino acid level. The 'bestfit' analysis, and an alignment with a low gap creation penalty, using the genetics computer group analysis package (Devereux *et al.*, 1984), suggests that the *tailless* sub-clone sequence aligns to a region of the *Drosophila melanogaster tailless* sequence in the second exon. Detailed comparison of this *Drosophila* region when aligned with *Xenopus* and chicken *tailless* gene sequences indicates that this region of the *tailless* gene has very low homology (less than 50%) between these different organisms. It is therefore very difficult to ascertain whether the level of homology seen between the *C. helgolandicus* putative *tailless* sequence and *Drosophila tailless* sequence is indicative of the *C. helgolandicus tailless* sequence being incorrect. The homology between the *Drosophila* sequence and the *C. helgolandicus tailless* sequence over the region of the sub-clone can be as high as 42%, but only by altering the guideline parameters of the computer program (reducing the gap creation penalty from 5 to 2 and the gap extension penalty from 0.3 to 0.1). Allowing the computer to create many gaps in the alignment in this way gives a rather ambiguous comparison between the two sequences. Such a low level of homology can make sequence analysis very difficult. For example, a direct pileup of chicken, *Xenopus* and *Drosophila tailless* sequences with the *C. helgolandicus tailless* sub-clone aligns the putative *C. helgolandicus tailless* gene completely erroneously. This alignment does not correspond to either the single alignment of *Drosophila* and *C. helgolandicus* sequences, the bestfit between *Drosophila* and *C. helgolandicus,* or to the region of *tailless* sequence originally used as the probe. Hence the identity of the *C. helgolandicus tailless* sub-clone remains ambiguous.

The approximately 400 bp band of the 207L λ clone *Xba*I digest was successfully sub-cloned and sequenced. Analysis of the 207L sub-clone sequence showed it to be

homologous to the *Antennapedia* class of homeobox genes. Comparisons of the 207L sequence at the nucleotide level with existing genes on the EMBL database are shown in Table 4.4.

Organism	% identity	Region of homology
Centipede (*Ethmostigmus ruripes*)	76.8 %	250 bp
Brine shrimp (*Artemia franciscana*)	75.5 %	220 bp
Spider (*Cupiennius salei*)	78 %	191 bp
Silkworm (*Bombyx mori*)	78.2 %	193 bp
Fruit fly (*Drosophila melanogaster*)	74.4 %	219 bp
Mosquito(*Anopheles gambiae*)	72.8 %	228 bp
Toad(*Xenopus laevis*)	79.4 %	189 bp
Mouse(*Mus musculis*)	64.4 %	354 bp

Table 4.4 Sequence identity of the *Calanus Antennapedia* gene at the nucleotide level compared to other homologous *Antennapedia* genes.

Comparisons of the 207L sequence at the amino acid level with existing genes on the EMBL database are shown in Table 4.5.

Organism	% homology	Region of homology
Centipede (*Ethmostigmus ruripes*)	92.9 %	70 aa
Brine shrimp (*Artemia franciscana*)	91.4 %	70 aa
Fruit fly (*Drosophila melanogaster*)	83.8 %	80 aa

Table 4.5 Sequence homology of the *Calanus Antennapedia* gene at the amino acid level compared to other homologous *Antennapedia* genes.

The *C. helgolandicus* 207L sequence is the first developmental gene to be found in *C. helgolandicus*. The homology it shares with other *Antennapedia* genes has prompted the 207L gene to be referred to as a *Calanus Antennapedia* gene (*Cal-Antp*). The sequence data for the *Cal-Antp* gene, in comparison to other *Antennapedia* homeobox genes, is shown in Figure 4.14.

```
                                                                  - - - - - - - - - - - - - -
Cal-Antp    TAA TCT AGA TTT ATT TTC TTT AAT AAA TTT CCT TTT TCA GAG AGG
Cal-Antp     *   S   R   F   I   F   F   N   K   F   P   F   S   E  >R
AfAntp                                                              >R
Dros Antp                                                           >R
Cen Antp                                                            >R

Cal-Antp    AAG CGA GGT CGC CAA ACC TAC ACT CGG TAC CAG ACT TTA GAA TTA
Cal-Antp     K   R   G   R   Q   T   Y   T   R   Y   Q   T   L   E   L
AfAntp       K   R   G   R   Q   T   Y   T   R   F   Q   T   L   E   L
Dros Antp    K   R   G   R   Q   T   Y   T   R   Y   Q   T   L   E   L
Cen Antp     K   R   G   R   Q   T   Y   T   R   Y   Q   T   L   E   L

Cal-Antp    GAA AAA GAA TTT CAC TTC AAC CGT TAC CTA ACT AGA AGA AGA AGA
Cal-Antp     E   K   E   F   H   F   N   R   Y   L   T   R   R   R   R
AfAntp       E   K   E   F   H   F   N   R   Y   L   T   R   R   R   R
Dros Antp    E   K   E   F   H   F   N   R   Y   L   T   R   R   R   R
Cen Antp     E   K   E   F   H   F   N   R   Y   L   T   R   R   R   R

Cal-Antp    ATA GAA ATC GCC CAC GCG CTA TGT CTG ACA GAA AGA CAA ATA AAA
Cal-Antp     I   E   I   A   H   A   L   C   L   T   E   R   Q   I   K
AfAntp       I   E   I   A   H   A   L   C   L   T   E   R   Q   I   K
Dros Antp    I   E   I   A   H   A   L   C   L   T   E   R   Q   I   K
Cen Antp     I   E   I   A   H   A   L   C   L   T   E   R   Q   I   K

Cal-Antp    ATA TGG TTT CAA AAC CGA AGG ATG AAG TGG AAA AAA GAG AAC AAA
Cal-Antp     I   W   F   Q   N   R   R   M   K   W   K   K   E   N   K
AfAntp       I   W   F   Q   N   R   R   M   K   W   K   K   E   N   K
Dros Antp    I   W   F   Q   N   R   R   M   K   W   K   K   E   N   K
Cen Antp     I   W   F   Q   N   R   R   M   K   W   K   K   E   N   K

Cal-Antp    TCA AAG CTG GAT GGA CCC GAC ATG GAC GAG TCA CCA GCG TCG AAC
Cal-Antp     S   K   L   D   G   P   D   M   D   E   S   P   E   S   N
AfAntp       S   K   V   D   S   G   C   L   D   G   I   L   V   D   H
Dros Antp    T   K   G   E   P   G   S   G   G   E   G   D   E   I   T
Cen Antp     A   K   L   E   G   A   G   G   D   L   C   L   T   G   L

Cal-Antp    TGA AAC TAA AAG TGA CTA ATG TTA GGG GTT CCA AAG AGT AGA GAG
Cal-Antp     *
AfAntp       V   L   G   M   *
Dros Antp    P   P   N   S   P   Q   *
Cen Antp     *
```

Figure 4.14 Sequence of the coding region of the 207L sub-clone, including additional 5' sequence obtained by IVPCR (Section 4.2.5). Translation of the 207L (*Cal-Antp*) sequence is shown with alignment to *Artemia franciscana* (*AfAntp*), *Drosophila melanogaster* (*Dros Antp*), and *Ethmostigmus ruripes* (*Cen Antp*) *antennapedia* genes. Putative intron sites are indicated with a >, putative intron sequence is shown in blue and the putative 3' splice acceptor site by a dashed overline. The homeodomain is denoted by red type. Stop codons = *

Figure 4.14 clearly illustrates the conserved homology between the homeodomains of *Antennapedia* genes from different organisms. Only the *Artemia franciscana* sequence (Averof and Akam, 1993) shows a single amino acid difference when the homeodomains of *Antennapedia* genes from *C. helgolandicus*, *Drosophila* (Averof and Akam, 1993), and a *centipede* (Grenier *et al.*, 1997) are compared. Each of these *Antennapedia* genes shares a common intron/exon boundary site directly preceding the homeodomain. The site of this 3' intron/exon boundary in the *C. helgolandicus Antennapedia* sequence reinforces the proposed identity of *Cal-Antp.*

Sequence similarities between the *Antennapedia* genes extend beyond the homeodomain, into downstream flanking regions. Out of the first 11 amino acids downstream of the homeodomain, *Cal-Antp* shares homology in 8 of the amino acids with at least 1 of the other *Antennapedia* sequences shown in Figure 4.14. This homology is consistent with that seen between homologous *Antennapedia* sequences from different organisms (Averof and Akam, 1993; Grenier *et al.*, 1997).

A stop codon can be seen for each sequence analysed, between 17 amino acids (*Cal-Antp* and Centipede *Antp*) and 23 amino acids (*Dros Antp*) downstream of the homeodomain. It is difficult to suggest whether this stop codon in *Cal-Antp* is representative of an intron/exon boundary site or of the termination of translation. Comparisons with homologous sequences suggest this stop codon could be either a putative splice donor site (*Artemia*, Averof and Akam, 1993; and centipede, Grenier *et al.*, 1997) or a putative translation stop (*Xenopus*, Fritz and De Robertis, 1988; Drosophila, Averof and Akam, 1993; and spider, Damen *et al.*, 1998).

These comparisons of *Cal-Antp* with homologous *Antennapedia* genes confirm that the sub-clone encodes a region of an *Antennapedia* protein homologue.

4.3.4 Analysis of expression of the *Calanus helgolandicus* developmental gene obtained from the library

4.3.4.1 PCR analysis of the solid phase cDNA library with one gene-specific primer and an oligonucleotide $(dT)_{25}$

Temporal expression of the *C. helgolandicus* developmental gene obtained from the genomic library, referred to as *Cal-Antp*, was analysed by amplification from the solid phase *C. helgolandicus* egg cDNA library with the gene-specific primer 207L-RTIb and an oligo $(dT)_{25}$. This amplification resulted in a band of approximately 600 bp when amplified with an annealing temperature of 40°C. Above this annealing temperature the resulting product became a smear. When capillary blotted onto a Hybond-N membrane and probed with the original clone 207 homeobox fragment and the 207L sub-clone sequence, the 207L-RTIa and oligo $(dT)_{25}$ band produced a strong hybridisation signal with both probes. The band was successfully cloned. However, sequencing showed the fragment to have the 207L-RTIa primer at both ends. Comparison with existing sequences on the EMBL database showed the fragment to be 67.4% homologous over a 175bp region to the gene encoding a human cleavage stimulation factor (CstF) (Takagaki and Manley, 1994).

It appeared impossible to analyse the temporal expression of the *Cal-Antp* gene by using one gene-specific primer and a oligo $(dT)_{25}$ reverse primer. This also prevented any cDNA sequence data outside the 207L sub-clone being obtained in this way.

4.3.4.2 PCR analysis of the solid phase cDNA library with two gene-specific primers

To analyse the expression of *Cal-Antp* the *C. helgolandicus* egg cDNA library was instead amplified with two gene-specific primers 207L-RTIb and 207L-RTII. The amplification product was a clear distinct band of the expected size (approximately 230 bp). Cloning and sequencing of this band confirmed the identity as being that of *Cal-Antp*.

The amplification was repeated on the solid phase cDNA libraries constructed from tissues of five different developmental stages. From this PCR analysis of the cDNA libraries it was evident that the *Antennapedia* gene encoded by the *Cal-Antp* sequence was expressed in a temporally specific manner. No expression was seen in eggs less than 18 hours old, but tissue taken from eggs greater than 18 hours old to copepodites showed significant expression. Expression appeared to cease in adult tissue (Figure 4.15).

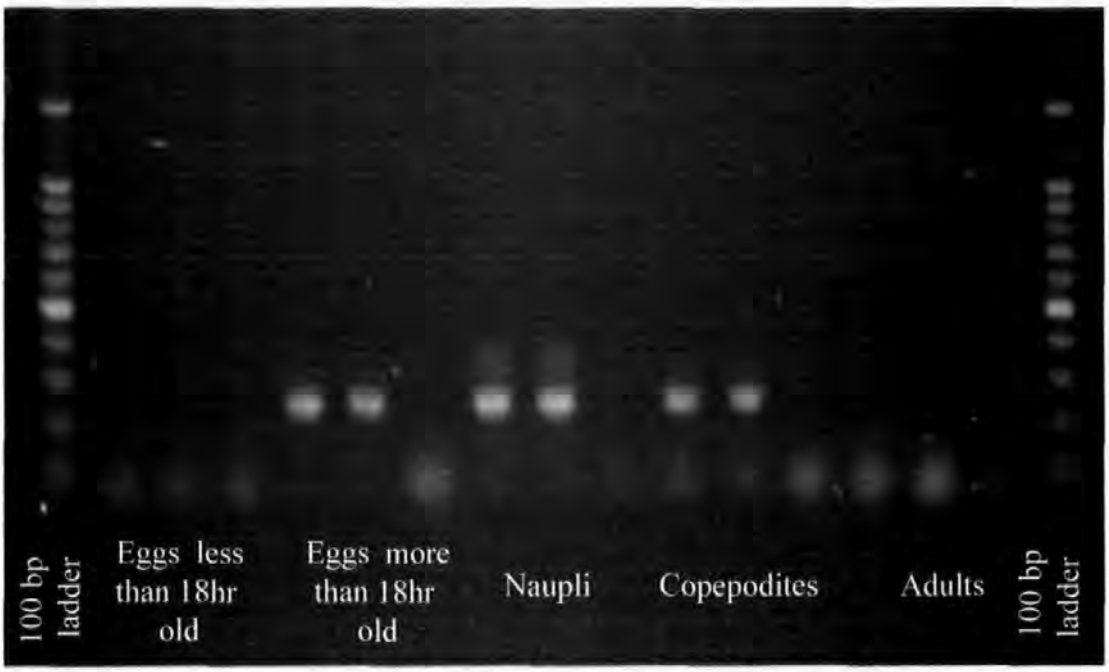


Figure 4.15 Expression of *Calanus Antennapedia* developmental gene (*Cal-Antp*). PCR analysis of 5 developmental stage cDNA libraries shows duration of expression. Each group of three lanes contains two replicates and a negative control.

4.3.5 Obtaining sequence data outside the 207L sub-clone by inverse PCR.

The original 207L λ clone restricted with *Dde*I and re-ligated to form circular molecules produced a strong PCR product when amplified solely with the first divergent primer pair 207L FI and RI (Figure 4.16). The second nested amplification with 207L F2 and R2 produced a less distinct PCR product and therefore it was the initial IVPCR band that was cloned and sequenced.

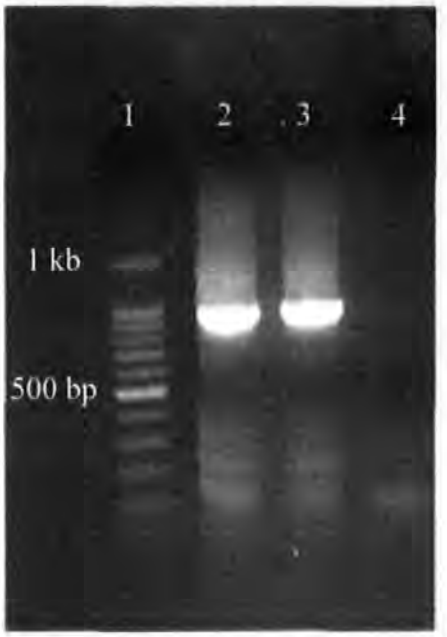


Figure 4.16 1% agarose gel showing the initial IVPCR amplification products of the *Dde*I restricted 207L λ clone with divergent primers designed to the 207L sub-clone. Lane 1 = 100bp DNA molecular weight marker; lanes 2 and 3 = 5 μL of the 207L IVPCR product; lane 4 = negative control.

The sequence of the IVPCR fragment was mapped to find the *Dde*I restriction site, and rearranged to give the contiguous genome sequence incorporating the original 207L sub-clone sequence. The total size of the sequenced fragment was 1171 bp. This provided sequence data upstream of the 207L sub-clone but within the putative intron (Figure 4.14). The extra sequence data provided by IVPCR downstream includes approximately 800 bp of non-coding sequence proceeding the stop codon seen in the original 207L sub-clone

4.4 Discussion

The *Calanus helgolandicus* library was initially screened with heterologous and homologous probes labelled with digoxigenin and detected by chemiluminescence. Unfortunately the results proved the inadequacies of the technique's ability to detect the DIG labelled probes when used to screen a large number of plaques. Instead, the *C. helgolandicus* library was screened with the two radioactively labelled heterologous gene probes, *tailless* and *even-skipped,* amplified from *Drosophila*, and the two radioactively labelled *C. helgolandicus* homologous gene probes clone 207 and clone 228. Primary, secondary and tertiary screens allowed the progressive isolation of individual plaques that hybridised to the *tailless* and clone 207 probes. Secondary screens of the plaques that initially hybridised to the *even-skipped* probe, clone 228 probe, and clone 207 probe at low stringency proved these plaques to be false positives. Considering that the screens with the *even-skipped*, clone 228, and clone 207 (low stringency) probes were on pre-screened membranes, stripped by treatment with 0.1% SDS and alkaline solution, or 0.1% SDS alone, it is fair to conclude that the stripping procedure was pernicious to the integrity of plaque DNA. This is somewhat surprising considering that the methods were tested in advance by probing, stripping and re-probing a membrane with a homologous 18S probe, with no hindrance to either the 2^{nd} or 3^{rd} probing. The manufacturer's instructions for the Hybond-N membranes also state that the membranes are suitable for multiple probing (Amersham Life Sciences). However, there is no other satisfactory explanation why these multiple probing attempts did not succeed. The fact that probing membranes, twice previously screened, at a low stringency with the clone 207 probe resulted in no positive hybridisation signals certainly reinforces the conclusion that the stripping procedure is pernicious to the plaques.

The sequence data of the sub-clone from the plaque that hybridised to the *tailless* probe, indicates a low level of homology with the *tailless* sequence. The original *tailless*

probe includes part of the conserved region of the *tailless* gene encoding the zinc finger protein. This region of the *tailless* gene shows considerable homology between *Drosophila*, *Xenopus* and chicken. However, the *C. helgolandicus tailless* sub-clone does not include any of the zinc finger coding sequence. The putative alignment of the *C. helgolandicus tailless* sub-clone sequence and *Drosophila tailless* gene is to a region of the gene showing little homology between different organisms, therefore making the alignment ambiguous. Presuming the *C. helgolandicus* sequence to be correct, a gene-specific primer was designed to this sequence and used in conjunction with a degenerate primer designed to the conserved region of the zinc finger protein. Amplification with these two primers on both the original λ *tailless* clone and *C. helgolandicus* genomic DNA failed to produce an amplification product. The clone has therefore not been pursued further and the identity remains uncertain.

The first unambiguous *C. helgolandicus* developmental gene (*Cal-Antp*) was obtained by screening the *C. helgolandicus* library with the homologous clone 207 *C. helgolandicus* homeobox probe. Sequencing of an isolated and sub-cloned fragment of this gene, plus further analysis by IVPCR, has shown it to be homologous with other *Antennapedia* homeobox genes.

The homeodomain of *Cal-Antp* shows complete homology with the centipede *Antennapedia* homeodomain (Grenier *et al.*, 1997), *Drosophila Antennapedia* domain (Averof and Akam, 1993), and 98% homology with *Artemia Antennapedia* homeodomain (Averof and Akam, 1993). Homology downstream of the homeobox, and putative intron/exon boundary sites also reinforce the proposed identity of *Cal-Antp*.

The *Antennapedia* class genes are homeotic selector genes. Homeotic selector genes are master genes that convert the repetitive pattern of segments into individual segments each with its own identity. They are coding factors for the antero-posterior axis of the body, not only in *Drosophila* but possibly in all animals (Slack, 1997). The

homeotic genes of *Drosophila* belong to two gene clusters called the *Antennapedia* complex (ANT-C) and the *Bithorax* complex (BX-C), jointly referred to as the homeotic complex HOM-C, which both lie on the right arm of the third chromosome (Lewis, 1978). The *Drosophila* homeotic genes closely resemble other invertebrate homeotic genes and also homeotic genes in vertebrates, referred to as *Antennapedia*-type or *Hox* genes. In vertebrates these *Hox* genes are arranged in four evolutionary related complexes, each on a different chromosome. Genes within each vertebrate *Hox* complex are aligned into 13 groups of genes called paralogues. This resemblance is not only structural, but also functional, with these genes in all animals being arranged along the chromosome as they are expressed along the antero-posterior axis (Gaunt *et al.*, 1988, Duboule and Dolle, 1989, Graham *et al.*, 1989).

The nomenclature for developmental genes can at best be confusing, the *Antennapedia* gene certainly being no exception, as *Antennapedia* is used to describe not only a single gene but a family, class and complex of genes. The actual *Antennapedia* gene is named after studies in *Drosophila* whereby a dominant transformation of antennae to mesothoracic leg occurred. In *Drosophila* it is suggested that the *Antennapedia* gene promotes leg development by repressing the antennal-determining genes, homeothorax and extradenticle (Cascares and Mann, 1998). During embryogenesis, in contrast with leg development, Antennapedia selects for a specific developmental pathway, determining larval mesothoracic pattern. This gene along with *Ultrabithorax*, *abdominal-A*, *fushi tarazu* and the vertebrate paralog groups (6, 7 and 8) form the *Antennapedia* family. Homeodomains within this family are highly conserved between insects and vertebrates (Bürglin, 1994). The *Antennapedia* family in turn is a member of the *Antennapedia* class of genes that are located in the centre of the Hom-C/*Hox* clusters.

It is assumed that there existed only 1 ancestral gene of the *Antennapedia* family, and that independent duplication events gave rise to paralog genes 6, 7 and 8 in

vertebrates, and to *Antennapedia*, *Abdominal-A*, *Ultrabithorax*, and *fushi tarazu* genes in flies (Akam, 1989; Kappen *et al.*, 1989; Krumlauf, 1992; Schubert *et al.*, 1993). Two of the middle body region-specifying genes, Ultrabithorax and Abdominal-A are unique to arthropods (Grenier *et al.*, 1997). Consideration of the theory of the evolution of the *Antennapedia* family explains the homology seen between *Cal-Antp* and homologous genes within this family. The *C. helgolandicus Antennapedia* gene directly corresponds to the *Antennapedia* genes of *Drosophila*, *Artemia* and centipede. It is known that all the homeotic genes that specify middle body regions in insects, including the *Antennapedia* family, originated before the divergence of the insect and crustacean lineages (Averof and Akam, 1993, Grenier *et al.*, 1997). Homology seen between *Cal-Antp* and vertebrate genes, such as the *Xenopus laevis Antennapedia* gene *XlHbox3* (*Hox A7*) and the human *Hox A7* and *Hox B7* genes, does not show a one to one correspondence but is indirectly analogous to the 7th vertebrate paralog group within the *Antennapedia* family.

The temporal expression of *Cal-Antp* was analysed through its mRNA complement. Possibilities of achieving such analysis include *in-situ* hybridisation, northern blot analysis and reverse transcription PCR (RT-PCR). Many reasons, including the limited amount of starting material available from *C. helgolandicus* developmental stages, led to the choice of RT-PCR. Analysis of expression of the *Cal-Antp* gene with a single gene-specific primer and an oligo (dT) primer proved to be ineffectual. However, amplification from cDNA with 2 gene-specific primers is well established, for example Damen *et al.* (1998). This technique proved effective in analysing the temporal expression of *Cal-Antp*. No expression was seen in eggs less than 18 h old or in adults, but *Cal-Antp* was expressed in tissue taken from eggs greater than 18h old and in nauplii and copepodite stages of *C. helgolandicus* development.

The duration of expression of *Cal-Antp* does not appear to render it ideal for use as an egg viability probe, with expression not being detected until eggs are over 18 hours old.

However, this is the first *C. helgolandicus* developmental gene to be isolated. Comparisons with existing gene sequences and analysis of expression of *Cal-Antp* have given us the first insight into the molecular mechanisms of *Calanus* development. Cloning and analysis of this *Antennapedia* homologue has given a perception of how the molecular mechanisms of *C. helgolandicus* development may relate to that of other arthropods and animals in general. The identification and analysis of *Cal-Antp* will also open new avenues in which to explore the reproductive and developmental biology of this important member of the zooplankton community.

CHAPTER 5

Identification and analysis of genes expressed during a defined developmental period

CHAPTER 5

Identification and analysis of genes expressed during a defined developmental period

5.1 Introduction

5.1.1 Identification of developmental genes expressed in *Calanus helgolandicus* eggs, via reverse transcription PCR analysis

5.1.2 Inverse PCR as a method of acquiring homeobox flanking sequences

5.2 Methods

5.2.1 Identification of developmental genes expressed in *Calanus helgolandicus* eggs via reverse transcription PCR analysis

5.2.1.1 Initial amplification of cDNA

5.2.1.2 Nested amplification

5.2.1.3 Cloning of homeobox fragments expressed in Calanus helgolandicus *eggs*

5.2.2 Inverse PCR analysis of homeobox flanking sequences

5.2.2.1 Amplification of homeobox flanking sequences by inverse PCR

5.2.2.2 Cloning of homeobox flanking sequences

5.3 Results

5.3.1 Reverse transcription-PCR analysis of developmental genes expressed in *Calanus helgolandicus* eggs.

5.3.1.1 Initial amplification of cDNA

5.3.1.2 Nested amplification

5.3.1.3 Sequence of homeobox fragments expressed in Calanus helgolandicus *eggs*

5.3.2 Inverse PCR analysis of homeobox flanking sequences

5.3.2.1 Amplified homeobox flanking sequences

5.3.2.2 Sequence of inverse PCR fragments

5.4 Discussion

5.1 Introduction

The cloning and analysis of specific *Calanus* developmental genes will advance our understanding of the mechanisms by which a *Calanus* egg develops into an adult. An understanding of the expression of such genes during development would enable investigation of the molecular basis of environmental effects on the developmental process. The expression analysis of such genes can also be applied to the improvement of existing marine ecology techniques, specifically the estimation of key recruitment processes such as copepod egg viability. A suitable developmental gene, expressed during early embryogenesis, will provide an excellent basis from which to develop a molecular probe to determine copepod egg viability. A critical factor in the development of such a molecular system lies with the correct expression period of the chosen developmental gene. The gene would need to be expressed post fertilisation, with continuous expression through early embryogenesis. Chapter 4 illustrated how specific developmental genes were acquired from a *C. helgolandicus* genomic library. In contrast, this chapter explains a second strategy for obtaining and analysing *C. helgolandicus* developmental genes, by targeting temporally specific expression.

The approach to probing a genomic library (Chapter 4) enables specific genes to be targeted, and has the advantage that the homologous gene used to probe the library will have been characterised for the organism from which it originates. However, this information can be applied only tentatively to *Calanus*, as there is no existing information which enables the *Calanus* developmental pathway to be compared with any other. This approach can be problematic as it relies on a suitable amount of homology being shared between the heterologous probe and the targeted *Calanus* developmental gene over a significant coding region.

In contrast, the method of obtaining genes by targeting time-specific expression will ensure that the gene is expressed during the desired period. As discussed in Chapter 4,

when homologous genes from different organisms are aligned it is sometimes possible to identify concise regions of conserved sequence identity. One such conserved region of developmental genes, the homeobox, has been identified. The homeobox is 180 bp in length, and codes for a highly conserved stretch of approximately 60 amino acids known as the homeodomain. By aligning homologous developmental genes degenerate oligonucleotide primers have been designed to the highly conserved regions (helices I and III) within the homeodomain. These degenerate primers have been used quite extensively to amplify a region of the homeobox from many different organisms, using the polymerase chain reaction. Time-specific expression of developmental genes can also be targeted via the highly conserved sequence motif of the homeobox using such degenerate primers.

5.1.1 Identification of developmental genes expressed in *Calanus helgolandicus* eggs, via reverse transcription PCR analysis

In addition to using the degenerate homeobox primers for the amplification of specific DNA sequences, these primers can also be used with RNA as the target material, provided that the RNA is first reverse transcribed to give a complementary strand of DNA (cDNA) to use as the template for amplification. Such a tissue-specific reverse transcription PCR strategy has successfully been used to acquire a novel murine homeobox gene (Kern *et al.*, 1992). This Chapter describes the application of a similar system to the recovery of *C. helgolandicus* homeobox-containing developmental genes expressed at a specific time of development.

5.1.2 Inverse PCR as a method of acquiring homeobox flanking sequences

When a small conserved region of a gene is the target site for amplification, as is the situation for homeobox amplification, a method of obtaining sequence data outside the conserved region is often necessary. The actual identity of homeobox clones acquired

through RT-PCR, which are limited to sequence data within the homeobox, can only tentatively be assigned through comparisons with other sequence similarities. For an accurate prediction it is necessary to acquire sequence information outside the homeobox region. This can be achieved using the technique known as Inverse PCR (IVPCR).

Standard PCR allows the amplification of DNA segments between two regions of known sequence. In 1988 a new procedure was designed which extends this technique to sequences that lie outside the boundaries of known sequences (Ochman *et al.*, 1988; Triglia *et al.*, 1988). This technique has undergone various modifications, but the basic principles remain (Silver, 1991). In brief, the technique involves restriction enzyme digestion of genomic DNA, and re-ligation of the DNA fragments, at a low concentration, to promote the formation of circular molecules. Specifically designed divergent primer pairs can then be used to amplify the flanking region of the known sequence (Figure 5.1).

IVPCR methods have been used to isolate genes of the *HOM/Hox* family from the crustacean *Artemia franciscana* (Averof and Akam, 1993). Recent knowledge was obtained on an improved IVPCR method which includes modifications for increased sensitivity (personal communication with Anna Goostrey, 1999). This method of IVPCR involves an initial amplification with primers designed to extend outwards from the homeobox. In contrast to standard PCR, the 3' ends of the hybridised primers should point away from each other. This is followed by a second, nested PCR with a new set of primers situated outside the first primers, still extending outwards from the centre of the homeobox. Such nested IVPCRs were performed on *C. helgolandicus* genomic DNA so as to acquire homeobox-flanking sequences.

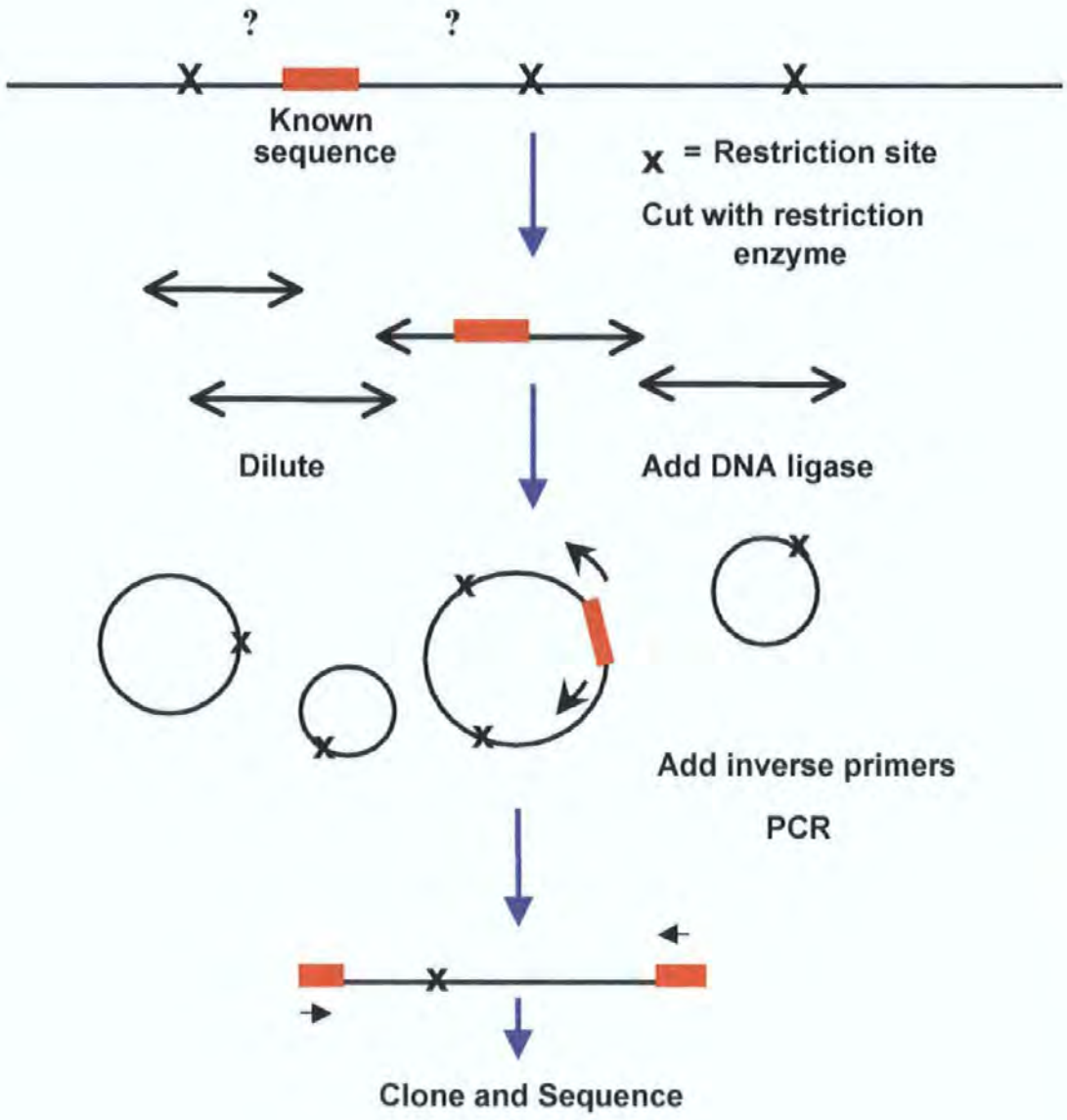


Figure 5.1 Diagrammatic representation of the basic principles involved in Inverse PCR (IVPCR).

5.2 Methods

5.2.1 Identification of developmental genes expressed in *Calanus helgolandicus* eggs, via reverse transcription PCR analysis

5.2.1.1 Initial amplification of cDNA

A solid phase cDNA library was constructed on oligonucleotide $(dT)_{17}$-coated Dynabeads representing the mRNA for all stages of *C. helgolandicus* eggs (Section 4.2.4). A small amount of this library was used for PCR analysis with the degenerate oligonucleotide homeobox forward primer, Hox C :-

```
                AC    A  A  A  C
5' - GCTCAGA--TNGA-AA-GA-TT-CA - 3'
                GT    G  G  G  T
```

and an oligonucleotide $(dT)_{17}$ reverse primer, complementary to the polyA$^+$ tail of the second strand cDNA.

An aliquot of approximately 50 μg of Dynabeads, representing a small aliquot of the cDNA library, was washed in 1 x *Taq* DNA polymerase buffer (10 μL) with the use of a magnetic separation stand (Promega UK Ltd). Prior to use, the *Taq* DNA polymerase buffer was removed and the beads re-suspended in 50 μL of the following PCR mix:-

Dynabeads with first strand cDNA	- (50 μg)
100 ng μL^{-1} Hox C primer	- 5 μL
100 ng μL^{-1} oligo $[dT]_{17}$ primer	- 5 μL
2 mM dNTP's	- 5 μL
10 x *Taq* DNA polymerase buffer	- 5 μL
Taq DNA polymerase	- 0.5 μL
MilliQ water	- 29.5 μL

A further 50 μL of the PCR mix provided a negative control, omitting any template DNA. Amplification was performed in a thermal cycler (PTC-100™, MJ Research, Inc.). The cycling parameters involved an initial cycle including a denaturation step of 94°C (1 min), an annealing step of 47°C (1 min) and an extension step of 72°C (5 min) to ensure

the complete synthesis of second strand cDNA. This initial cycle was followed by a denaturation step at 94°C (4 min) during which the Dynabeads and attached first strand cDNA were removed. The amplification was then continued with 35 cycles of 47°C (1 min), 72°C (1.5 min) and 94°C (1 min). A final annealing phase at 47°C (2 min) was followed by an extension phase at 72°C (5 min) and storage at 4°C until use.

The amplified products were analysed by 1% agarose gel electrophoresis.

5.2.1.2 Nested amplification

Due to the low level of specificity achievable during the first round of amplification it is frequently desirable to re-amplify the initial products with specific nested primers. A small portion of the initial reaction was therefore re-amplified with the original Hox C and oligonucleotide $(dT)_{17}$ primers, and also with a specifically designed rcHox D primer (reverse complement of Hox D) and oligonucleotide $(dT)_{17}$ primer. Both combinations were applied to maximise the potential for obtaining a precise band of a homeobox developmental gene, which by the nature of the primers would contain sequence information flanking the homeobox. A third combination of primers was used, that of the original Hox C and Hox D degenerate homeobox primers.

The re-amplifications were performed on 1 µL of template DNA with the following PCR mix.

Template DNA	- 1 µL
Forward primer	- 5 µL
Reverse primer	- 5 µL
2 mM dNTP's	- 5 µL
Taq DNA polymerase 10 x buffer	- 5 µL
Taq DNA polymerase	- 0.5 µL
MilliQ water	- 28.5 µL

Amplifications were performed in a thermal cycler (PTC-100^{TM}, MJ Research, Inc.). The cycling parameters involved an initial denaturation step of 94°C (1 min), followed by 35 cycles of 47°C (1 min), 72°C (1 min) and 94°C (2 min). A final annealing step of 47°C (2 min) was followed by a long extension step at 72°C (5 min) and storage at 4°C. The resulting products were analysed by 1 % agarose gel electrophoresis.

5.2.1.3 Cloning of homeobox fragments expressed in Calanus helgolandicus *eggs*

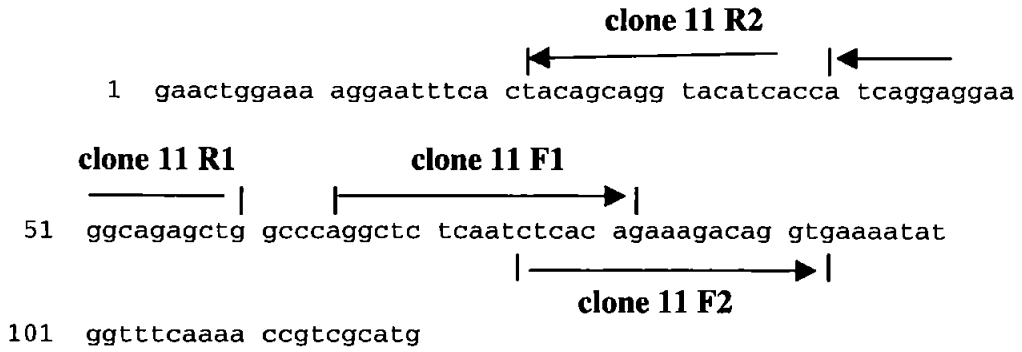
The re-amplification product from the Hox C and Hox D primer reaction, subsequent to agarose gel electrophoresis (1%), was recovered using a Qiaex II extraction kit (Section 2.3.7). The recovered amplified fragments were then cloned into pBluescript SK$^-$, following the methods described in Sections 2.3.8 – 2.3.10 and 2.2.6. The recombinant clones were identified by colony PCR (Section 2.3.11) and sequencing was achieved on seven of the clones, following the chain-termination method described in Sections 2.3.13 - 2.3.17. An aliquot of transformed *E. coli* for each of the different recombinant clones was prepared for long-term storage following the method in Section 2.2.9. Additionally, recombinant plasmid DNA was recovered from an *E. coli* culture (for each unique clone) using Promega's Wizard miniprep DNA purification system (Section 2.3.12), and the preparation stored long term at –20°C.

5.2.2 Inverse PCR analysis of homeobox flanking sequences

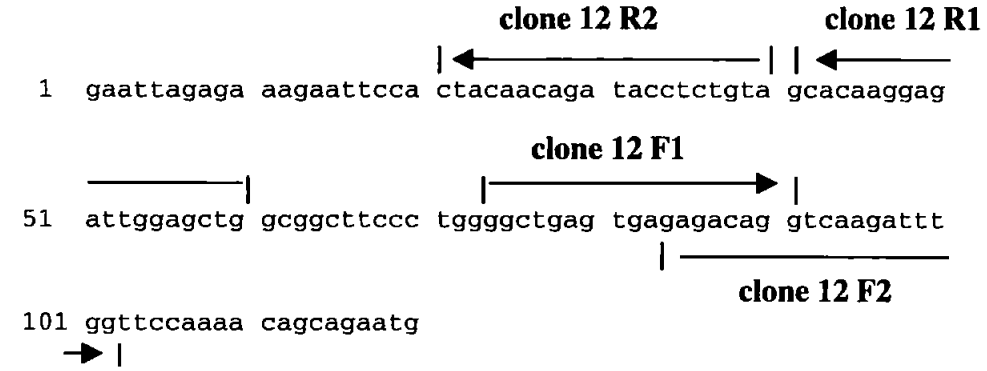
5.2.2.1 Amplification of homeobox flanking sequences by inverse PCR

Two specific pairs of primers were designed for each of four cloned and sequenced homeobox fragments, targeted in *C. helgolandicus* eggs. The primers were designed with as little overlap as possible, and with careful consideration of length, melting temperature and self-hybridization. The following figure illustrates the position of each primer on the four homeobox clones analysed.

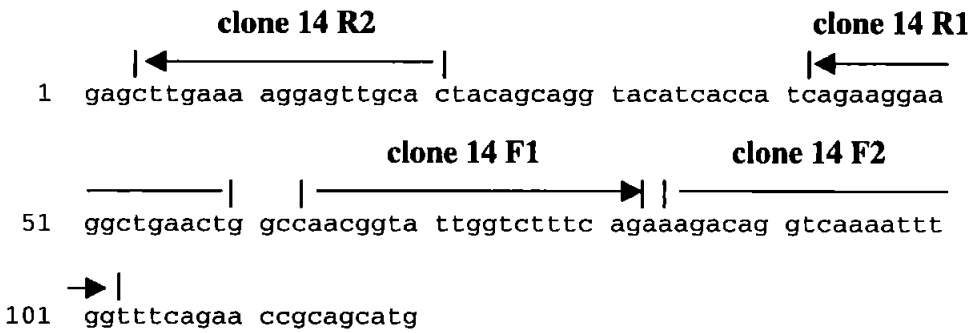
Homeobox Clone 11



Homeobox Clone 12



Homeobox clone 14



Homeobox clone 20

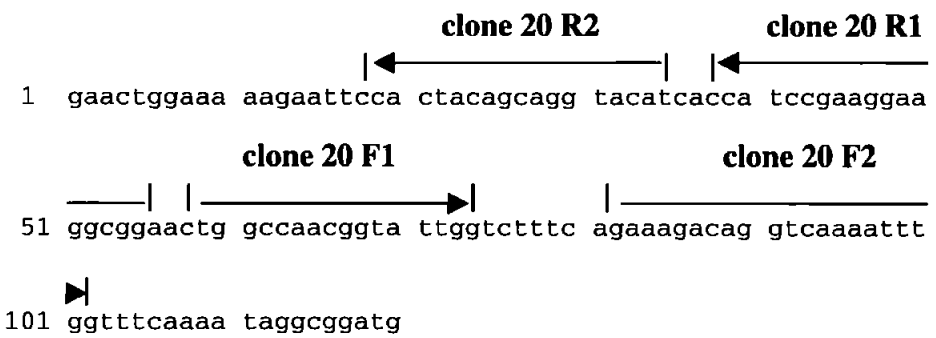


Figure 5.2 IVPCR primers designed for the four homeobox clones 11, 12, 14 and 20.

These clones are referred to as clone 11, clone 12, clone 14, and clone 20.

For the IVPCR reactions a large-scale preparation of DNA from 500 *C. helgolandicus* was undertaken following the method described in Section 3.2.2.1. The quality of the extracted DNA and a rough estimation of the quantity were checked by electrophoresis through a low percentage agarose gel.

The original homeobox sequences were mapped to determine which of the restriction enzymes would target a recognition site inside the homeobox regions, and would therefore not be suitable. A range of 4-bp cutters and 6-bp cutters, with which to digest the *C. helgolandicus* genomic DNA, were then chosen. Two of the chosen restriction enzymes, *Dde*I and *Eco*R1, were shown to cut within the homeobox region of clone 12, and one enzyme, *Eco*RI, within the homeobox region of clone 20. These particular reactions were therefore not subjected to amplification with the corresponding primers. Aliquots of approximately 1 µg of DNA were digested with 10 enzymes (Table 5.1), in the following reaction mixture:-

Genomic DNA	- 1 µL
10 x enzyme reaction buffer	- 1 µL
1 mg mL^{-1} BSA	- 1 µL
Enzyme	- 1 µL
MilliQ water	- 6 µL

The reaction mixtures were carefully mixed with a pipette tip and then incubated for 3.5 h at the correct temperature for each respective enzyme (Table 5.1).

Enzyme	Size of target site (bp)	Nature of ends produced	Optimum enzyme temperature
*Xba*I	6	sticky	37°C
*Eco*RI	6	sticky	37°C
*Dde*I	4	sticky	37°C
*Vsp*I	6	sticky	37°C
*Hin*dIII	6	sticky	37°C
*Sma*I	6	blunt	25°C
*Eco*RV	6	blunt	37°C
*Sau*3AI	4	sticky	37°C
*Bam*HI	6	sticky	37°C
*Hha*I	4	sticky	37°C

Table 5.1 Nature of enzymes used to digest *Calanus helgolandicus* genomic DNA.

Following incubation, each reaction was made up to a volume of 400 µL with MilliQ water and then extracted once with an equal volume of phenol/chloroform:isoamyl alcohol (24:1), and once with chloroform:isoamyl alcohol (24:1). The aqueous phase was precipitated with 1/10 volume NaOAC and 2.5 volumes 100% ethanol at –80°C overnight. The resulting washed and dried pellet was re-suspended in 23 µL of 10 x ligase buffer and 205 µL MilliQ water.

Ligations were performed at a DNA concentration of approximately 2.8 ng μL^{-1} to favour the construction of circular molecules as opposed to concatamers. To each re-suspended pellet, 8 units of DNA ligase (2 µL) and rATP to a final concentration of 0.8 mM (20 µL) were added. The ligations were incubated at 16°C for 20 h, followed by heat inactivation of the ligase at 70°C for 10 min. The circularised genomic DNA was stored at 4°C as a stock for use in the IVPCR reactions.

Initial IVPCR reactions were set up with the four sets of homeobox primers (clone 11 F1 and R1, clone 12 F1 and R1, clone 14 F1 and R1, and clone 20 F1 and R1) for all ten of the circularised DNA reactions, except where the restriction enzyme had been found to

cut within the homeobox region. The IVPCR reaction mixture included the following amounts of reactants:-

Aliquot of circularised DNA	- 2 μL
(100 ng μL^{-1}) Primer F1	- 2.5 μL
(100 ng μL^{-1}) Primer R1	- 2.5 μL
10 x *Taq* DNA polymerase buffer	- 2.5 μL
25 mM MgCl	- 1.0 μL
2 mM dNTP's	- 2.5 μL
Taq DNA polymerase	- 0.5 μL
Single stranded binding protein (2.2 mg mL^{-1})	- 0.08 μL
MillQ water	- 11.42 μL

Amplification was performed in a thermal cycler (PTC-100™, MJ Research, Inc.). The cycling parameters involved an initial denaturation step of 94°C (4 min), followed by 35 cycles of 53°C (30 sec), 72°C (1 min) and 94°C (30 sec). A final annealing step of 53°C (1 min) was followed by a long extension step at 72°C (5 min) and storage at 4°C. A 1 μL aliquot of the PCR product was then used directly in the second, nested IVPCR, using the corresponding F2 and R2 primers. The reaction mix was as follows:-

Initial IVPCR product	- 1.0 μL
(100 ng μL^{-1}) Primer F2	- 2.5 μL
(100 ng μL^{-1}) Primer R2	- 2.5 μL
10 x *Taq* DNA polymerase buffer	- 2.5 μL
25 mM MgCl	- 1.0 μL
2 mM dNTP's	- 2.5 μL
Taq DNA polymerase	- 0.5 μL
MillQ water	- 12.5 μL

Amplification was performed using the same parameters as for the initial IVPCR. 10 μL of each PCR product were analysed by 0.8 % agarose gel. Of the 37 nested IVPCR amplifications performed, four indicated a positive amplification product. These potentially positive reactions were repeated in triplicate using 1 μL of the initial PCR product. A negative control, lacking template DNA, was included.

5.2.2.2 Cloning of homeobox flanking sequences

For each of the positive IVPCR amplifications 26 μL were separated by 1% agarose gel electrophoresis, and the resulting bands recovered using a Qiaex II gel extraction kit (Section 2.3.7). The DNA was treated with *Pfu* DNA polymerase to create blunt termini (Section 2.3.9) and ligated into prepared pBluescript SK^- (Section 2.3.10). The recombinant plasmids were transformed into chemically competent *E. coli* cells (Section 2.2.6) and grown on LB agar ampicillin plates, with the addition of X-gal and IPTG (Section 2.2.8). Recombinant colonies were identified by colony PCR (Section 2.3.11). The corresponding recombinant plasmid DNA was then recovered from *E. coli* using Promega's Wizard miniprep DNA purification system (Section 2.3.12) and the transformed *E. coli* cells stored long-term at –80°C (Section 2.2.9). 50 μL of each plasmid preparation were ethanol precipitated (Section 2.3.5) at –20°C overnight. The pellets were washed with 70 % ethanol, vacuum dried, and the DNA sequences determined commercially by MWG Biotech (UK) limited (guaranteed accuracy of 99%).

The sequences were mapped to locate the position of the original restriction sites then rearranged to give the contiguous genome sequence. This involved incorporating the central region of homeobox sequence which exists between the IVPCR primers. The sequences are shown in Figures 5.7, 5.8, and 5.9. The corresponding amino acid sequences are illustrated in Figures 5.7a, 5.8a, and 5.9a.

5.3 Results

5.3.1 Reverse transcription-PCR analysis of developmental genes expressed in *Calanus helgolandicus* eggs

5.3.1.1 Initial amplification of cDNA

The results of the initial amplification of cDNA from *C. helgolandicus* eggs with the degenerate homeobox primer Hox C and a reverse primer complimentary to the polyA$^+$ tail of the mRNA are shown in Figure 5.3.

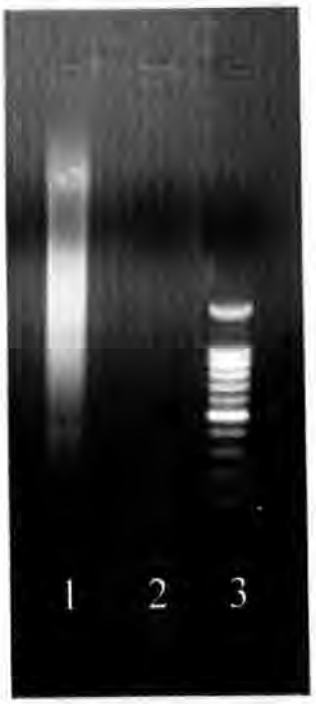


Figure 5.3 Agarose gel (1%) showing the RT-PCR product amplified with an oligonucleotide homeobox forward primer (Hox C) and an oligonucleotide $(dT)_{17}$ reverse primer. Lane 1 = 5 μL of amplified PCR product as described; lane 2 = 5 μL of negative control; and lane 3 = 5 μL of 100 bp DNA marker.

As Figure 5.3 illustrates, the product of the PCR is a smear containing many amplified fragments of variable length. It is therefore not possible directly to isolate and clone any single gene from within this smear. To obtain a defined developmental gene sequence this PCR product was therefore subjected to re-amplification with different primer combinations in nested amplification reactions.

5.3.1.2 Nested amplification

The results from the re-amplification, with various primer combinations, of the initial Hox C and oligonucleotide $(dT)_{17}$ PCR product, are shown in Figure 5.4.

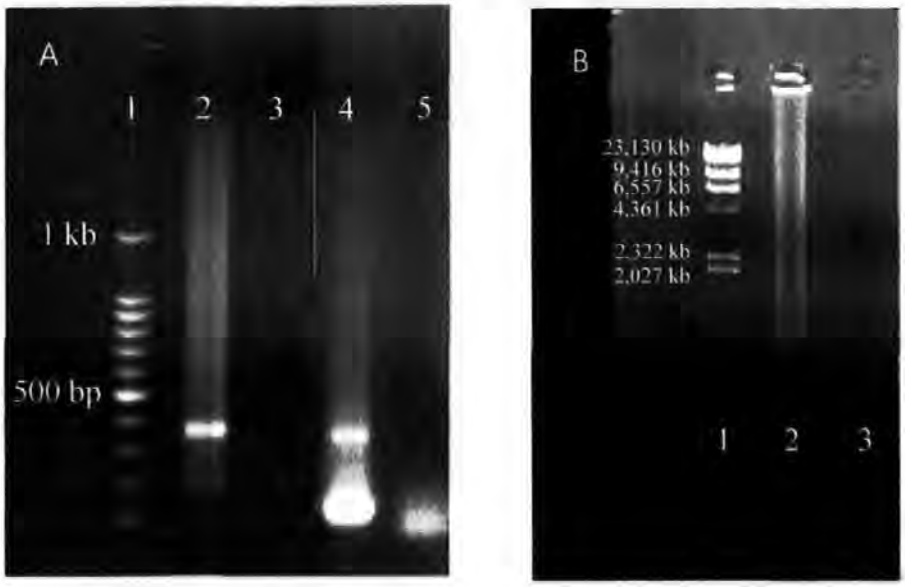


Figure 5.4 Agarose gels (1%) showing the re-amplification PCR products of the initial Hox C/oligonucleotide smear (Figure 5.3) with various primer combinations. Gel A:- lane 1 = 4 μL 100 bp DNA marker; lane 2 = 5 μL of PCR product re-amplified with Hox C and oligonucleotide $(dT)_{17}$ primers; lane 3 = negative control; lane 4 = 5 μL of PCR product re-amplified with Hox C and Hox D primers; lane 5 = negative control. Gel B:- lane 1 = 200 ng λ *Hin*dIII DNA marker; lane 2 = 5 μL of PCR product re-amplified with a rcHox D primer and oligonucleotide $(dT)_{17}$ primer; lane 3 = negative control.

The re-amplification with the original Hox C and oligonucleotide $(dT)_{17}$ primers resulted in a strong band of approximately 380 bp in length (Figure 5.4, gel A, lane 2). Further analysis of this PCR product however, proved the product to be a result of non-specific hybridisation. The re-amplification with the specifically designed rcHox D primer and oligonucleotide $(dT)_{17}$ primer resulted in a smear, similar to that of the original amplification (Figure 5.4, gel B, lane 2). Direct isolation and cloning of any particular DNA fragment is impossible from such a range of PCR products. In contrast, the re-amplification with the two homeobox primers, Hox C and Hox D, did produce a clear bright band of the expected size (120 bp) (Figure 5.4, gel A, lane 4). A much fainter band,

exactly the same size as the Hox C/oligonucleotide $(dT)_{17}$ re-amplification band was also visible within this same lane.

The re-amplification with Hox C and Hox D primers appeared suitable for further analysis (Section 5.2.1.3).

5.3.1.3 Sequence of homeobox fragments expressed in Calanus helgolandicus *eggs*

The band re-amplified from the original Hox C/oligonucleotide $(dT)_{17}$ PCR product with the Hox C and Hox D primers was further analysed by cloning and sequencing. This band, although of a concise size, contained many homeobox sequences. This is to be expected given the nature of the degenerate primers.

Of the seven recombinant clones sequenced there were two pairs of identical sequences, the sequences are shown in Figure 5.5.

```
Clone 11+17  gaactggaaa aggaatttca ctatagcagg tacatcacca tcaggaggaa
Clone 12     gaattagaga aagaattcca ctacaacaga tacctctgta gcacaaggag
Clone 13     gaatttgaga aggagtttca tttctttcat tttctgacag gagttcgtag
Clone 14+15  gagcttgaaa aggagttgca ctacagcagg tacatcacca tcagaaggaa
Clone 20     gaactggaaa aagaattcca ctacagcagg tacatcacca tccgaaggaa

Clone 11+17  ggcagagctg gcccaggctc tcaatctcac agaaagacag gtgaaaatat
Clone 12     attggagctg gcggcttccc tggggctgag tgagagacag gtcaagattt
Clone 13     gactgaaatg gcccacacac ttggattaac agagaggcag ataaaaattt
Clone 14+15  ggctgaactg gccaacggta ttggtctttc agaaagacag gtcaaaattt
Clone 20     ggcggaactg gccaacggta ttggtctttc agaaagacag gtcaaaattt

Clone 11+17  ggtttcaaaa ccgtcgcatg
Clone 12     ggttccaaaa cagcagaatg
Clone 13     ggttccaaaa tcgacgtatg
Clone 14+15  ggtttcagaa ccgcagcatg
Clone 20     ggtttcaaaa taggcggatg
```

Figure 5.5 Sequence data for seven cloned homeobox fragments, expressed in *Calanus helgolandicus* eggs. The primers Hox C (start of sequence) and Hox D (end of sequence) are denoted by bold type.

By comparing these fragments of homeobox sequences with sequences in the EMBL database it was possible to obtain some idea of the class of each putative homeobox gene as follows:-

Clone 11 (and 17) showed 80 % homology over 114 bp with the *Xenopus laevis Xcad-3* gene, 78 % homology over 113 bp with the mouse *cdx4-a* gene and 72 % homology over 120 bp with a chicken *caudal* type gene.

Clone 12 showed 74 % homology over 117 bp with the *Xenopus laevis Xcad-1* gene, and over the whole region 73 % homology with *Heliocidaris erythrogramma* (sea urchin) *HEHBOX9* gene, and 74 % homology with the *Xenopus laevis XlHbox-8* gene.

Clone 13 showed 70 % homology over 115 bp with *Schistosoma* (an early bilateral metazoan) *Antennapedia* gene, and 71 % over the whole region with the mouse *Hox-3.5* gene.

Clones 14 and 15 were identical and showed 79 % homology over 113 bp with the chicken homeobox protein *Cdx-3*, 78 % homology over 116 bp with a human *Cdx-2* gene, and 75 % homology over the same region with a carp *Cdx1* gene.

Clone 20 showed 77 % homology over 116 bp with a silkworm homeobox gene *caudal* homologue, 74 % homology over 113 bp with the mouse *Cdx-4* gene, and 70 % homology over the whole region with a chicken *HoxD-3* gene.

These comparisons with known sequences from the EMBL database suggest that clones 11, 17, 14, 15 and 20 appear to belong to the *caudal* class of homeobox genes, and clone 13 to the *Antennapedia* complex. Additional sequence data outside of these homeobox regions was needed to establish further the identity of the putative homeobox genes. Sequence data of the flanking regions was acquired through IVPCR.

5.3.2 Inverse PCR analysis

5.3.2.1 Amplified homeobox flanking sequences

Of the 37 nested IVPCR amplifications performed, four showed a distinct amplification product as follows: clone 11 primers used on the *Hha*I digested and circularised DNA; clone 12 primers used on the *Sau*3AI digested and circularised DNA; and clone 20 primers used on both the *Sau*3AI and *Hha*I digested and circularised DNA. These potentially positive reactions were repeated in triplicate and the results are shown in Figure 5.6.

Figure 5.6 Agarose gel (1%) showing positive IVPCR amplified products. Lane 1 = 4 µL 100 bp DNA marker; lanes 2, 3 and 4 = *Sau*3AI digested and circularised DNA amplified with clone 12 primers; lane 5 = negative control; lanes 6, 7, and 8 = *Sau*3AI digested and circularised DNA amplified with clone 20 primers; lane 9 = negative control; lanes 10, 11 and 12 = *Hha*I digested and circularised DNA amplified with clone 11 primers; lane 13 = negative control; lanes 14, 15, and 16 = *Hha*I digested and circularised DNA amplified with clone 20 primers; lane 17 = negative control; lane 18 = 4 µL 100 bp DNA marker.

For each of the four reactions the replicate samples produced clear, distinct products of the same size, with no products apparent in the negative controls.

5.3.2.2 Sequence of inverse PCR fragments

The sequences of the homeobox containing developmental gene fragments identified from *C. helgolandicus* eggs through RT-PCR and IVPCR have been rearranged to give the contiguous genome sequence, incorporating the central region of homeobox sequence that exists between the IVPCR primers. These sequences are referred to as clone 11-IVPCR, clone 12-IVPCR and clone 20-IVPCR. The sequences are shown in Figures 5.7, 5.8, and 5.9. The corresponding putative amino acid sequences are illustrated in Figures 5.7a, 5.8a, and 5.9a.

Having obtained sequence data flanking the homeobox regions, it is possible to make a very accurate prediction of the identity of the putative homeobox genes. Both the nucleotide sequence data and the translated amino acid data were used to make comparisons with other sequences, using the genetics computer group (GCG) sequence analysis software package (Devereux *et al.*, 1984). The three putative homeobox gene sequences were used to interrogate the Genbank and EMBL databases with the following results:

Clone 11-IVPCR sequence: Best matches at the nucleotide level are 76% identity (over 161 bp) with *Anopheles gambiae* (malaria mosquito) *caudal* gene, 78% identity (over 144 bp) with chicken *cHox-cad* gene, 76% identity (over 153 bp) with mouse *cdx-2* homeobox gene, 74% identity (over 164 bp) with the human homeobox protein *cdx2,* and others. At the amino acid level the best matches are 78% identity (over 89 aa) with *Bombyx mori* (silkmoth) *caudal* protein, 86% identity (over 72 aa) with *Tribolium castaneum* (red flour beetle) *caudal* protein and 91% identity (over 65 aa) with chick homeobox protein *cHox-cad.*

Clone 12-IVPCR sequence: Some of the best matches at the nucleotide level are 71% identity (over 187 bp) with chicken *HoxB3* gene, 70 % identity (over 212 bp) with human *Hox2I* gene, and 66% identity (over 223 bp) with *Drosophila melanogaster Scr* (Sex

combs reduced) homeotic gene. At the amino acid level the homology stretches considerably outside the homeodomain region. The best matches are 43.9% identity (over 168 aa) with chicken *HoxB-3 gene*, 43.6% (over 169 aa) with human *HoxA3* gene, and 38% (over 200 aa) with Amphioxus *Hox3* gene.

Clone 20-IVPCR Sequence: Some of the best matches at the nucleotide level are 72% identity (over 252 bp) with *Bombyx mori caudal* gene, 76% (over 207 bp) with chicken *Cdx-C caudal* gene, and 75% (over 199 bp) with human *Cdx-2 caudal* gene. At the amino acid level the best matches are 82.1% (over 78 aa) with *Bombyx mori caudal* gene, 81.3% (over 75 aa) with *Tribolium castaneum caudal* gene, 85.5% (over 69aa) with chicken *cHox-cad* gene, and 84% (over 69 aa) with *Xenopus laevis Xcad* 2 gene.

Figure 5.7 shows 686 bp of a *C. helgolandicus caudal* type homeobox containing developmental gene (clone 11-IVPCR). Interestingly this sequence contains an intron situated within the homeobox. The intron is localised 132 bp from the beginning of the homeobox sequence thus interrupting the homeodomain sequence between amino acids 44 and 45. The length of the intron is 205 bp and it is flanked by the sequences 5'-AGGTGATT –3' and 5'- TGCAGG –3' representing possible donor and acceptor sites for the splicing of the primary RNA transcripts. These splicing junctions conform to the GT-AG rule (Shapiro *et al.*, 1987).

Detailed comparison between the chicken *caudal* gene homologue (*CHox-cad*) (Frumkin *et al.*, 1991), the *Bombyx caudal* gene homologue (Xu *et al.*, 1994), *Drosophila caudal* gene (*cad*) (Mlodzik and Gehring, 1987) and the putative *C. helgolandicus caudal* gene homologue (clone 11-IVPCR) showed that these protein products are most homologous in the region of the homeodomains. *C. helgolandicus* and *Bombyx* are 93% homologous in the homeodomain region, *C. helgolandicus* and chicken are 92% homologous and *C. helgolandicus* and *Drosophila* 90%. It is known that this homology in

caudal proteins can extend upstream of the homeodomain to the hexapeptide sequence (Frumkin *et al.*, 1991). This homology can not be seen when compared to *C. helgolandicus* as the sequence data extends only 2 amino acids upstream of the homeodomain (Figure 5.7a). However these 2 amino acids preceding the homeobox are conserved between *C. helgolandicus*, *Drosophila*, *Bombyx*, chicken, mouse and others. A strong homology between *C. helgolandicus* and *Bombyx caudal* proteins can also be seen in the first few amino acids following the homeodomain. Of the first 13 amino acids following the homeodomain, 8 are identical and 3 show conservative changes.

The stop codon of the *C. helgolandicus caudal* protein in frame with the homeobox is present 40 amino acids downstream of the homeodomain. This compares with the stop codon that marks the end of the *cHox-cad* protein, which is present 51 amino acids downstream of the homeodomain, and the stop codon in *Bombyx caudal* protein which is present 65 amino acids downstream of the homeodomain.

Figure 5.7 Re-arrangement of the IVPCR sequence amplified with clone 11 primers. The central homeobox sequence has been incorporated to give the contiguous genome sequence. The positions of the nested IVPCR primers are shown. Base pairs in red denote the homeobox region, and base pairs shown in blue indicate the position of the intron. 5' and 3' putative splice sites are shown as dashed overlined.

Clone 11	R	T	**K**	**D**	**K**	**Y**	**R**	**V**	**V**	**Y**	**S**	**D**	**H**	**Q**	**R**	**L**	**E**	**L**	**E**	**K**
CHox-cad			**K**	**D**	**K**	**Y**	**R**	**V**	**V**	**Y**	**T**	**D**	**H**	**Q**	**R**	**L**	**E**	**L**	**E**	**K**
Bombyx-cad			**K**	**D**	**K**	**Y**	**R**	**V**	**V**	**Y**	**S**	**D**	**H**	**Q**	**R**	**L**	**E**	**L**	**E**	**K**

Clone 11	**E**	**F**	**H**	**Y**	**S**	**R**	**Y**	**I**	**T**	**I**	**R**	**R**	**K**	**A**	**E**	**L**	**A**	**Q**	**A**	**L**
CHox-cad	**E**	**F**	**H**	**Y**	**S**	**R**	**Y**	**I**	**T**	**I**	**R**	**R**	**K**	**A**	**E**	**L**	**A**	**A**	**A**	**L**
Bombyx-cad	**E**	**F**	**H**	**Y**	**S**	**R**	**Y**	**I**	**T**	**I**	**R**	**R**	**K**	**A**	**E**	**L**	**A**	**V**	**S**	**L**

Clone 11	**N**	**L**	**T**	**E**	**R**	**Q↔V**		**K**	**I**	**W**	**F**	**Q**	**N**	**R**	**R**	**A**	**K**	**E**	**R**	**K**
CHox-cad	**G**	**L**	**T**	**E**	**R**	**Q**	**V**	**K**	**I**	**W**	**F**	**Q**	**N**	**R**	**R**	**A**	**K**	**E**	**R**	**K**
Bombyx-cad	**G**	**L**	**S**	**E**	**R**	**Q**	**V**	**K**	**I**	**W**	**F**	**Q**	**N**	**R**	**R**	**A**	**K**	**E**	**R**	**K**

Clone 11	**Q**	**V**	K	K	R	E	E	L	I	H	K	D	K	I	D	L	P	G	I	A
CHox-cad	**V**	**N**																		
Bombyx-cad	**Q**	**V**																		

Clone 11	V	H	P	G	H	V	A	A	M	S	G	K	E	I	T	K	Q	L	Q	E

Clone 11	I	N	*

Figure 5.7a Translation of the clone 11-IVPCR sequence, starting from nucleotide 1 and terminating at the first stop codon. The intron shown in Figure 5.7 has been removed and the site of removal indicated by ↔. The clone 11-IVPCR protein sequence is aligned with the homeodomains from the chicken *caudal* protein (*CHox-cad*) and *Bombyx mori caudal* protein (*Bombyx-cad*) for comparison. The homeodomain region is shown in bold type.

Figure 5.8 shows a 777 bp fragment of a putative *C. helgolandicus* developmental gene (clone 12-IVPCR) showing strong homology to the *Antennapedia* class. This clone 12-IVPCR sequence shows 67% identity within the homedomain to the chick *HoxB3* gene (Scotting *et al.*, 1990; Rex and Scotting, 1994). A further 15 amino acids show a conservative change (Figure 5.8a). Upstream of the homeodomain in chick *HoxB3* is a hexapeptide, common to many developmental genes. A putative hexapeptide can be seen in the clone 12-IVPCR sequence (underlined in Figure 5.8a), which has three identical and two conservatively changed amino acids when compared with chick *HoxB3*.

Clone 12-IVPCR also shows 69% identity within the homeodomain, to the amphioxus homeobox gene, *AmphiHox3*, (Holland *et al.*, 1992). A further 14 amino acids show a conservative change. Comparison of the hexapeptide upstream of the homeodomain in *AmphiHox3* reveals that clone 12-IVPCR has 4 identical amino acids and two that are conservatively changed. A third region of conservation in the *HoxB3* genes is at the N-terminus. No comparison with the clone 12-IVPCR sequence however, can be made at this location, as the sequence does not continue far enough upstream.

The third gene, for which there is available information, which shows a strong homology to clone 12-IVPCR, is the *Drosophila Sex Combs Reduced* gene (*Src*). This gene is also a member of the *Antennapedia* class of homeotic genes. At the protein level the sequence identity within the homeodomain is 65% with clone 12-IVPCR. However, at the nucleotide level the gene shows 66% homology over 223 bp, an area obviously extending outside the homeobox. In the hexapeptide region clone 12-IVPCR again shares four identical amino acids and two conservatively changed amino acids with *Scr*.

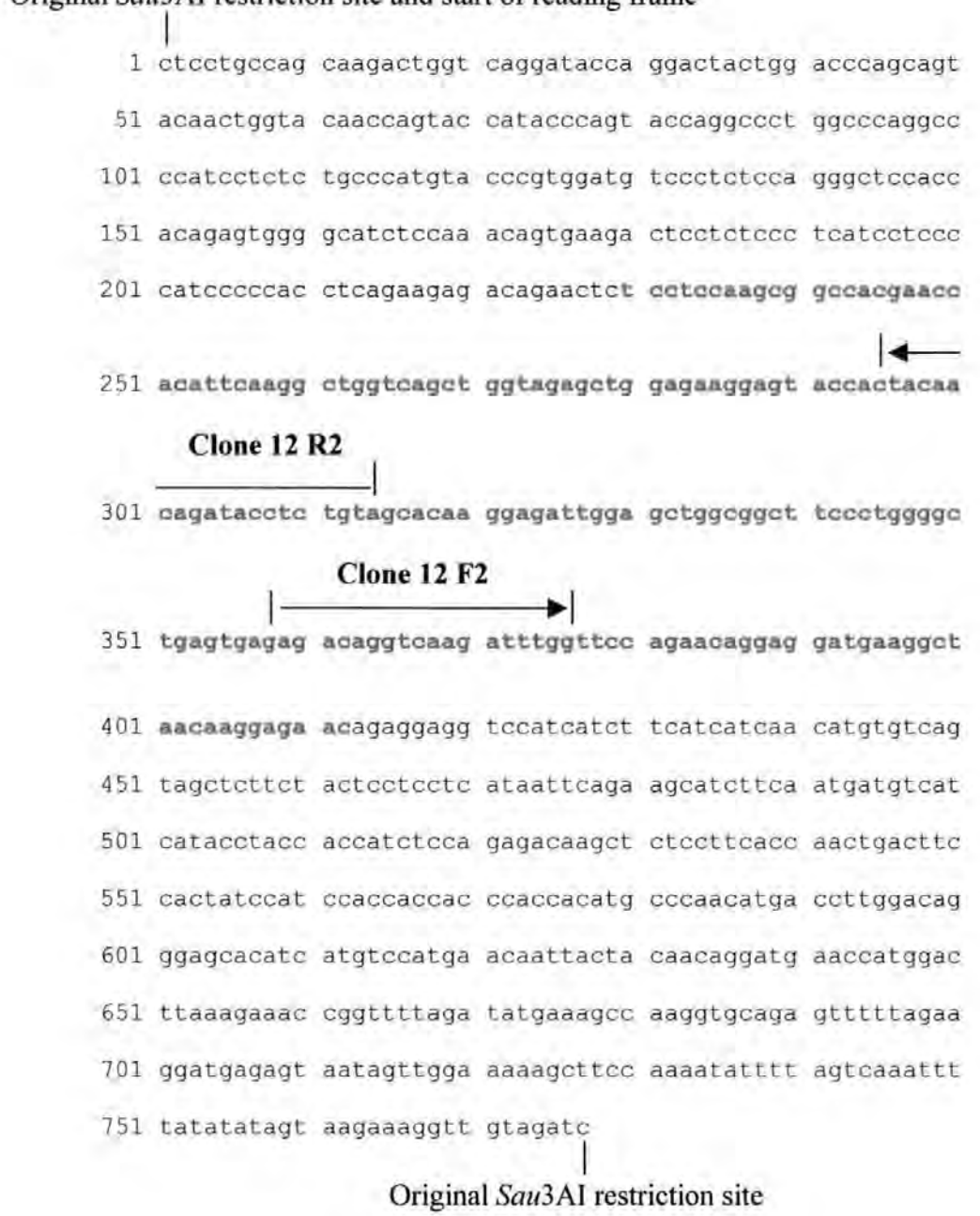


Figure 5.8 Re-arrangement of the IVPCR sequence amplified with clone 12 primers. The central homeobox sequence has been incorporated to give the contiguous genome sequence. The positions of the nested IVPCR primers are shown. Base pairs in red denote the homeobox region.

Clone 12	S	C	Q	Q	D	W	S	G	Y	Q	D	Y	W	T	Q	Q	Y	N	W	Y
Clone 12	N	Q	Y	Q	T	Q	Y	Q	A	L	A	Q	A	P	S	S	L	P	M	Y
Clone 12	P	W	M	S	L	S	R	A	P	P	Q	S	G	A	S	P	N	S	E	D
Clone 12	S	S	P	S	S	S	P	S	P	T	S	E	E	T	E	L	S	**S**	**K**	**R**
Scr																		**T**	**K**	**R**
ChHoxB3																		**S**	**K**	**R**
AmpHox3																		**G**	**K**	**R**
Clone 12	**P**	**R**	**T**	**T**	**F**	**K**	**A**	**G**	**Q**	**L**	**V**	**E**	**L**	**E**	**K**	**E**	**Y**	**H**	**Y**	**N**
Scr	**Q**	**R**	**T**	**S**	**Y**	**T**	**R**	**Y**	**Q**	**T**	**L**	**E**	**L**	**E**	**K**	**E**	**F**	**H**	**F**	**N**
ChHoxB3	**A**	**R**	**T**	**A**	**Y**	**T**	**S**	**A**	**Q**	**L**	**V**	**E**	**L**	**E**	**K**	**E**	**F**	**H**	**F**	**N**
AmpHox3	**A**	**R**	**T**	**A**	**Y**	**T**	**S**	**A**	**Q**	**L**	**V**	**E**	**L**	**E**	**K**	**E**	**F**	**H**	**F**	**N**
Clone 12	**R**	**Y**	**L**	**C**	**R**	**P**	**R**	**R**	**L**	**E**	**L**	**A**	**A**	**S**	**L**	**G**	**L**	**S**	**E**	**R**
Scr	**R**	**Y**	**L**	**T**	**R**	**R**	**R**	**R**	**I**	**E**	**I**	**A**	**H**	**A**	**L**	**C**	**L**	**T**	**E**	**R**
ChHoxB3	**R**	**Y**	**L**	**C**	**R**	**P**	**R**	**R**	**V**	**E**	**M**	**A**	**N**	**L**	**L**	**N**	**L**	**S**	**E**	**R**
AmpHox3	**R**	**Y**	**L**	**C**	**R**	**P**	**R**	**R**	**V**	**E**	**M**	**A**	**A**	**M**	**L**	**N**	**L**	**T**	**E**	**R**
Clone 12	**Q**	**V**	**K**	**I**	**W**	**F**	**Q**	**N**	**R**	**R**	**M**	**K**	**A**	**N**	**K**	**E**	**N**	**R**	G	G
Scr	**Q**	**I**	**K**	**I**	**W**	**F**	**Q**	**N**	**R**	**R**	**M**	**K**	**W**	**K**	**K**	**E**	**H**	**K**		
ChHoxB3	**Q**	**I**	**K**	**I**	**W**	**F**	**Q**	**N**	**R**	**R**	**M**	**K**	**Y**	**K**	**K**	**D**	**Q**	**K**		
AmpHox3	**Q**	**I**	**K**	**I**	**W**	**F**	**Q**	**N**	**R**	**R**	**M**	**K**	**Y**	**K**	**K**	**E**	**Q**	**K**		
Clone 12	P	S	S	S	S	S	T	C	V	S	S	S	S	T	P	P	H	N	S	E
Clone 12	A	S	S	M	M	S	S	Y	L	P	P	S	P	E	T	S	S	P	S	P
Clone 12	T	D	F	H	Y	P	S	T	T	T	H	H	M	P	N	M	T	L	D	R
Clone 12	E	H	I	M	S	M	N	N	Y	Y	N	R	M	N	H	G	L	K	E	T
Clone 12	G	F	R	Y	E	S	Q	G	A	E	F	L	E	G	*					

Figure 5.8a Translation of the clone 12 IVPCR sequence, starting from nucleotide 1 and terminating at the first stop codon. Three *Antennapedia* class gene homeodomains are shown for comparison, Chick *HoxB3* (*ChHoxB3*), Amphioxus *Hox3* gene (*AmpHox3*) and *Drosohphila Sex Combs Reduced* (*Scr*). The homeobox is denoted in bold type. A putative hexapeptide sequence is underlined.

Figure 5.9 shows a 411 bp fragment of a *C. helgolandicus caudal* type homeobox-containing developmental gene (clone 20-IVPCR). Figure 5.9a shows part of the putative translation sequence of this gene, including the homeodomain.

Detailed comparisons of *caudal* homeodomains show clone 20-IVPCR to have 90% identity with *Bombyx* (Xu *et al.*, 1994) and 87% identity with the red flour beetle (*Tribolium castaneum*) (Schulz *et al.*, 1998). Additionally the clone 20-IVPCR sequence shows 87% identity with the clone 11-IVPCR homeodomain and 80% identity within the region that the clone 20-IVPCR and clone 11-IVPCR sequences overlap (75 aa). Unlike the clone 11-IVPCR sequence, clone 20-IVPCR contains no introns inside the homeobox. However, from a detailed comparison of the clone 20-IVPCR sequence with other *caudal* gene sequences it would appear that there is an intron present, 5 amino acids upstream of the homeobox. This is supported by a putative 3' splicing site in the correct location, shown in Figure 5.9. Comparison of the 5 amino acids following the putative intron and preceding the homeodomain, show that *C. helgolandicus* (clone 20-IVPCR) has 4 identical amino acids and 1 conservatively changed amino acid, when compared with the same region in chicken, *Bombyx, Tribolium, Xenopus* and other *caudal* genes.

Downstream of the homeobox the first 11 amino acids of the clone 20-IVPCR sequence, when compared with both the *Bombyx* and *Tribolium caudal* amino acid sequences, show 5 amino acids to be identical and 5 to have been conservatively changed. The sequence data for clone 20-IVPCR does not extend as far as a stop codon, so no comparisons to homologous genes regarding the length proceeding the homeobox can be made.

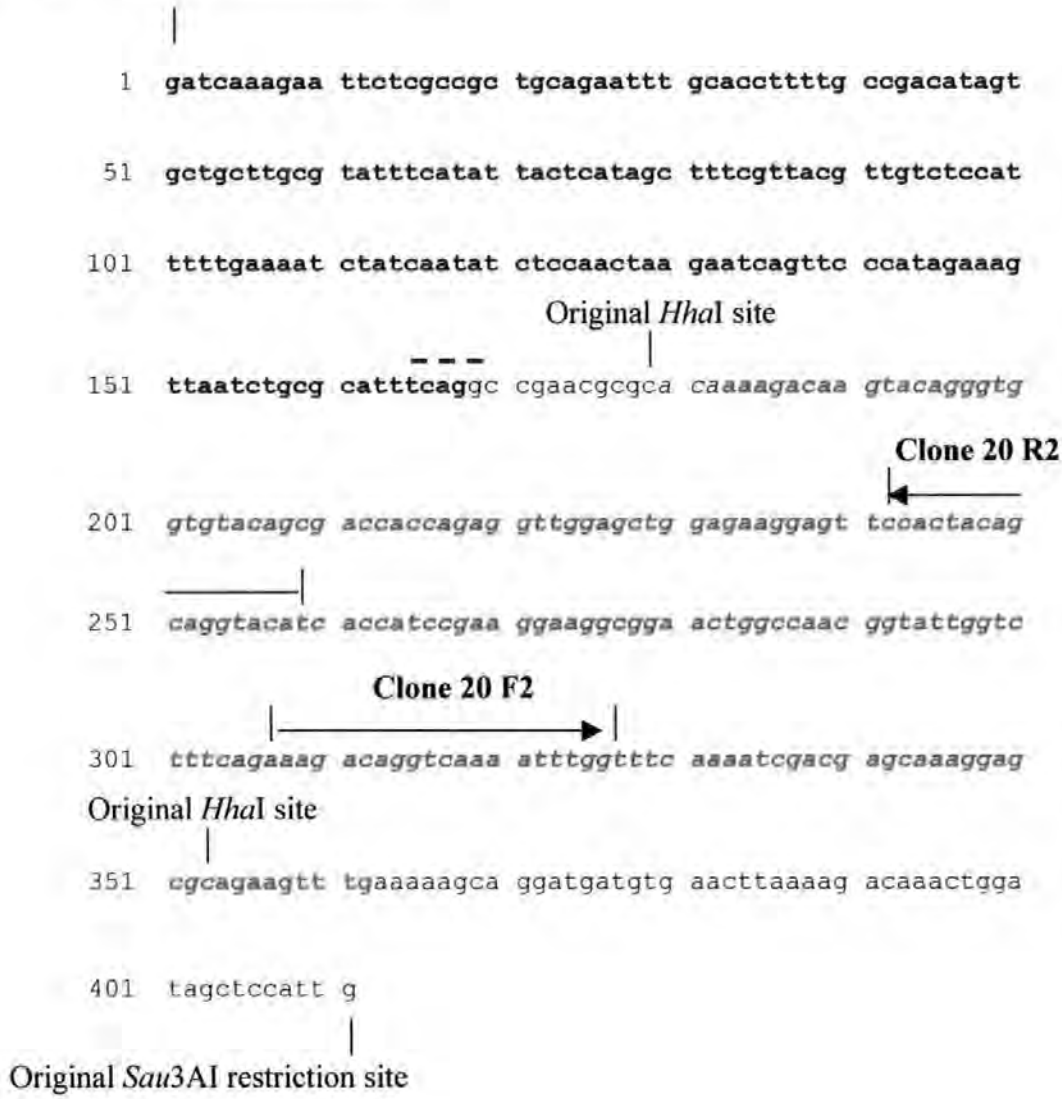


Figure 5.9 Re-arrangement of the IVPCR sequence amplified, from circularised *Sau*3AI digested *C. helgolandicus* DNA, with clone 20 primers. The re-arrangment of the IVPCR sequence, from circularised *Hha*I digested *C. helgolandicus* DNA, also amplified with the clone 20 primers, is identical and therefore contained within this longer sequence (position indicated by italics). The central homeobox sequence has been incorporated to give the contiguous genome sequence. The positions of the nested IVPCR primers are shown. Base pairs in red denote the homeobox region, and base pairs shown in blue indicate the position of a putative intron. 3' putative splice site is shown as dashed overlined.

Clone 20	*	K	V	N	L	R	I	S	G	R	T	R	T	K	D	K	Y	R	V	V	Y
Clone 11												R	T	K	D	K	Y	R	V	V	Y
Bombyx-cad														K	D	K	Y	R	V	V	Y
Beetle-*cad*														K	D	K	Y	R	V	V	Y

Clone 20	S	D	H	Q	R	L	E	L	E	K	E	F	H	Y	S	R	Y	I	T	I	R
Clone 11	S	D	H	Q	R	L	E	L	E	K	E	F	H	Y	S	R	T	I	T	I	R
Bombyx-cad	S	D	H	Q	R	L	E	L	E	K	E	F	H	Y	S	R	Y	I	T	I	R
Beetle-*cad*	T	D	H	Q	R	V	E	L	E	K	E	F	Y	Y	S	R	Y	I	T	I	R

Clone 20	R	K	A	E	L	A	N	G	I	G	L	S	E	R	Q	V	K	I	W	F	Q
Clone 11	R	K	A	E	L	A	Q	A	L	N	L	T	E	R	Q	V	K	I	W	F	Q
Bombyx-cad	R	K	A	E	L	A	V	S	L	G	L	S	E	R	Q	V	K	I	W	F	Q
Beetle-*cad*	R	K	A	E	L	A	N	S	L	G	L	S	E	R	Q	V	K	I	W	F	Q

Clone 20	N	R	R	A	K	E	R	R	S	L	K	K	Q	D	D	V	N	L	K	D	K
Clone 11	N	R	R	A	K	E	R	K	Q	V	K	K	R	E	E	L	I	H	K	D	K
Bombyx-cad	N	R	R	A	K	E	R	K	Q	V											
Beetle-*cad*	N	R	R	A	K	E	R	K	Q	V											

Clone 20	L	D	S	S	I
Clone 11	I	D	L	P	G

Figure 5.9a Translation of the clone 20 IVPCR sequence, starting from the first stop codon and terminating at the end of the cloned sequence. The clone 11-IVPCR translated sequence and the homeodomains from *caudal* gene homologues of *Tribolium castaneum* (beetle-*cad*) and *Bombyx mori* (*Bombyx-cad*) are shown for comparison. The amino acids in blue denote the putative intron and the homeodomains are indicated by bold type.

5.4 Discussion

Three homeobox containing developmental genes have been successfully identified and analysed, through their expression in *C. helgolandicus* embryos. These genes have the potential to provide powerful molecular tools with which to estimate the key recruitment factor involving copepod egg viability.

The initial identification of developmental genes by RT-PCR analysis with the Hox C and oligo-$(dT)_{17}$ primers resulted in a smear of various length amplification products, preventing the isolation of any single gene fragment. This PCR product was therefore subjected to re-amplification with different primer combinations to maximise the potential to obtain a defined amplified product of a homeobox containing developmental gene. As the results show, re-amplification with rcHox D and oligo-$(dT)_{17}$ primers again produced a smear of amplification products, probably because a number of homeobox containing developmental genes were amplified with the original degenerate Hox C and oligo-$(dT)_{17}$ primers. Re-amplification with rcHox D and oligo-$(dT)_{17}$ primers again amplified a number of different homeobox-containing genes of various lengths. It is known that the polyA$^+$ tail can be hundreds of base pairs in length (Wormington, 1993; Curtis *et al.*, 1995). Therefore, to eliminate the possibility that the smear was a result of the oligo-$(dT)_{17}$ primer annealing to various places along the polyA$^+$ tail, both the original and re-amplifications were performed with an anchored oligo-$(dT)_{17}$ primer. The anchored oligo-$(dT)_{17}$ primer was designed with 17 T's plus two additional bases at the 3' end. The penultimate base was a degeneracy of either A, C or G, and the last base an N. This should have ensured that the anchored oligo-$(dT)_{17}$ primer annealed to the 5' end of the polyA$^+$ tail. The anchored oligo-$(dT)_{17}$ primer made no difference to the amplification results, confirming that the smears were a result of many homeobox containing genes being amplified and not purely an artefact of the oligo-$(dT)_{17}$ primer annealing to various places along the polyA$^+$ tail. Re-amplification of the original Hox C and oligo-$(dT)_{17}$ smear with

the same primers produced a positive product which required further analysis. Such analysis was achieved by performing a re-amplification with the Hox C and Hox D primers to establish if the DNA fragments represented in the band contained a homeobox. Only a weak band of the appropriate size could be seen when the products were run on an agarose gel. A single primer control PCR amplification was performed using only the Hox C primer, and unfortunately this produced a band of the same size, indicating that the PCR product was being amplified by non-specific hybridisation of the Hox C primer. Cloning and sequencing of the PCR product confirmed the presence of the Hox C primer on both ends of the DNA fragment. No match with any developmental genes could be established when aligning the DNA sequence with existing sequences on the EMBL database. Therefore the band was dismissed as the product of non-specific primer annealing.

The third primer combination (Hox C and Hox D) used to re-amplify the original Hox C and oligo-$(dT)_{17}$ smear was the least suitable option, as only sequence information inside the homeobox could be obtained. However, it proved the most successful, with the amplification of a number of homeoboxes from developmental genes expressed in *C. helgolandicus* embryos. The fainter band also seen in this amplification (380 bp in size) was further analysed and found to be a non-specific annealing product with Hox C at both ends. The cloning of the distinct 120 bp band produced by re-amplification of the original Hox C and oligo-$(dT)_{17}$ smear with Hox C and Hox D primers resulted in the discovery of seven homeobox fragments expressed in *C. helgolandicus* eggs. However, because the homeobox is highly conserved and of a small size, sequence data outside the homeobox region was necessary to determine accurately the identity of the developmental gene from which the homeobox belongs. Since the method of RT-PCR has proven inadequate to obtain sequence data outside the homeobox regions as any PCR with a homeobox specific primer and oligo-$(dT)_{17}$ reverse primer results in a smear of amplification products, IVPCR was used as a method of obtaining sequence data flanking the homeoboxes. The IVPCR

successfully provided sequence data flanking three homeoboxes, these putative developmental gene sequences being referred to as clone 11-IVPCR, clone 12-IVPCR and clone 20-IVPCR.

Comparisons of these gene sequences with existing gene sequences on the computer database have given a good idea of the identification and role of these putative *C. helgolandicus* developmental genes. Clone 11-IVPCR and Clone 20-IVPCR both represent homeobox containing developmental gene fragments expressed in *C. helgolandicus* embryos. The homeodomains of these developmental genes are homologous to the family of *caudal* genes. At both the nucleotide and amino acid level they show a strong identity with *caudal* genes from organisms ranging from the silkmoth (Xu *et al.*, 1994) and the red flour beetle (Schulz *et al.*, 1998) to the mouse (James *et al.*, 1994) and humans (Mallo *et al.*, unpublished). Within the homeodomain the *C. helgolandicus* clone 11-IVPCR *caudal* homologue shows greater than 90% identity with silkmoth, red flour beetle, chicken, *Drosophila* and other *caudal* homeodomains. This is a significantly high homology. To compare, the *Bombyx caudal* and *Drosophila cad* homeodomains are 80% identical (Xu *et al.*, 1994), and the *Chox-cad* shows 80.3% identity with *Drosophila cad* homeodomain and 95% with mouse *cdx1* homeodomain (Frumkin *et al.*, 1991). In comparison the homeodomain of the *C. helgolandicus* clone 20-IVPCR *caudal* homologue shows an identity of 90% or below, with *caudal* homeodomains from other organisms.

The role of *caudal* genes, based on the knowledge of the *Drosophila* developmental pathway, has been shown to be involved in the initial establishment of morphogenic gradients. Unlike most of the other developmentally significant genes *caudal* (*cad*) was not identified initially from a mutant phenotype, but was cloned because it contained a homeobox (Slack, 1991). The *caudal* gene is also unusual in that until recently, although a homeobox gene, it was not thought to be a member of a known homeobox gene cluster.

Recent studies however, have shown that the amphioxus *caudal* gene is a member of a novel homeobox gene cluster (ParaHox), thus challenging the idea that *caudal* is a 'dispersed' *hox*-like gene (Brooke *et al.*, 1998).

By comparison with other *caudal* gene proteins it would be fair to assume that both the clone 11-IVPCR and clone 20-IVPCR sequences represent only part of putative *Calanus caudal* genes. The putative protein sequence, omitting non-coding sequence, obtained by RT-PCR and IVPCR extends to 100 amino acids in length for clone 11-IVPCR, and 78 amino acids in length for clone 20-IVPCR. In comparison *Bombyx caudal* gene encodes a protein of 244 amino acids (Xu *et al.*, 1994) and the chicken *caudal* transcript is 248 amino acids long (Frumkin *et al.*, 1991). No sequence homologous to a consensus start sequence can be seen in either clone 11-IVPCR or clone 20-IVPCR.

Interestingly, the homeodomain of clone 11-IVPCR is interrupted by an intron. By comparison of the clone 11-IVPCR homeobox cDNA sequence obtained by RT-PCR, and the equivalent genomic sequence obtained by IVPCR, as well as the sites of putative splicing regions, the intron can be accurately located between amino acids 44 and 45 of the homeodomain. The homeodomain sequence of the *CHox-cad* protein is also broken at exactly the same point between amino acids 44 and 45 of the homeodomain (Frumkin *et al.*, 1991).

The presence of an intron that interrupts the homeodomain is a relatively rare observation, particularly in vertebrate homeobox genes. In *Drosophila* however, a number of homeoboxes have been identified in which the sequences are interrupted by an intron. A common place for this to occur in *Drosophila* is between amino acids 44 and 45, with such a location for introns in *Drosophila* homeobox genes being seen in *Labial* (Mlodziz *et al.*, 1988), *Abdominal-B* (DeLorenzi *et al.*, 1988), *Distal-less* (Cohen *et al*; 1989) and *NK-1* (Kim and Nirenberg, 1989). Therefore, the intron in clone 11-IVPCR is in a position that is not uncommon for fly homeobox genes.

Introns have also been found in *caudal* genes from other organisms, such as *Bombyx, Drosophila*, and chicken, but these are located upstream of the homeobox. Such an intron can not be seen in the clone 11-IVPCR sequence because the sequence data does not extend this far upstream of the homeobox. However, in the clone 20-IVPCR sequence it would appear that there is an intron, 5 amino acids upstream of the homeodomain. The location of this intron is identical to the location of introns found in the *caudal* genes of *Bombyx* (Xu *et al.*, 1994), chick (Frumkin *et al.*, 1992) and *Drosophila* (Mlodzik and Gehring, 1987). The putative boundary of the intron-exon can be localised in clone 20-IVPCR by the position of a splice acceptor site. It is interesting to note that the homology upstream from the homeobox in various *caudal* genes ceases at the position where the gene is interrupted by the intron. The precise conservation of these intron-exon boundary positions reinforces the proposed homologies of clone 11-IVPCR and clone 20-IVPCR.

Caudal represents a key developmental gene in a wide range of organisms and it would be fair to assume that the *C. helgolandicus caudal* gene homologues play a key role in the development of *C. helgolandicus*. Such a gene could be used to monitor the effects of environmental factors on the expression of key developmental genes. It is also possible that one of the genes may be suitable for use as an egg viability probe. Further analysis of clone 11-IVPCR and clone 20-IVPCR temporal expression in developing embryos is therefore of great importance. Although there is no substitute for direct analysis we can gain a brief insight into their possible expression by comparison with the expression of homologous genes.

In *Drosophila* the *cad* gene encodes maternal and zygotic transcripts that are derived from different promoters. *Caudal* maternal RNA is regulated at the level of translation by the *bicoid* homeodomain and *cis*-acting sequences in the 3' untranslated region of the *caudal* message (Dubnau and Struhl. 1996). The negative regulation by *bicoid* results in the *caudal* protein accumulating along the anterior-posterior axis in a

gradient complementary to the anterior-posterior *bicoid* gradient. This gradient is replaced at later stages of development by the zygotic RNA and protein which, after gastrulation is localised to the posterior midgut and Malpighian tubules, the posterior midgut being of endodermal origin (Mlodzik and Gehring, 1987). The *Bombyx caudal* gene is also a maternal gene, with a single maternal transcript of 2.3 kb. The *Bombyx caudal* protein forms gradients along the anteroposterior axis during gastrulation, restricted to the most posterior domain two days after embryogenesis (Xu *et al.*, 1994). In contrast, the *cad* gene in ascidians, which is responsible for tail formation, is expressed only in the zygote (Katsuyama, 1999). The *Cdx1* homeobox gene, during the course of murine development, is also expressed only post fertilisation, in embryos 14 days *post coitum* (p.c.). After this stage its transcripts are localised to the epithelial lining of the intestine, which is of endodermal origin (Duprey *et al.*, 1988). As with the mouse *caudal* gene the chicken *CHox-cad* gene is not a maternal gene but is expressed between 16 hours and 4 days of chick embryo development. The onset of *CHox-cad* transcription correlates with the onset of gastrulation, suggesting a role for this gene during the establishment of the three germ layers (Frumkin *et al.*, 1991). The *Xenopus laevis caudal* type gene is expressed also at the beginning of gastrulation and remains a constant level through the end of neurulation (Duboule, 1994; Pownall *et al.*, 1996).

It is known that clone 11-IVPCR and clone 20-IVPCR are expressed in the early stages of *C. helgolandicus* embryogenesis as they were targeted by time specific expression in *C. helgolandicus* embryos. However, it is most important to establish whether the *C. helgolandicus caudal* gene homologues are maternally expressed, as is the case for *Drosophila* and *Bombyx*. Or, as with the other organisms discussed, whether *C. helgolandicus caudal* gene homologues are zygotic genes involved in endodermal development. A thorough understanding of clone 11-RTPCR and clone 20-RTPCR

expression in *C. helgolandicus* developing embryos is needed before their use as an egg viability probe can be considered.

It also interesting to note the presence of two *caudal* type genes within an invertebrate such as *Calanus*. Although unlikely, it is possible that one of the putative *caudal* gene homologues from *C. helgolandicus* originates from a different homeobox gene, containing a remarkably similar homeobox sequence. However, the homology with the homeobox flanking sequences in other *caudal* genes, including the positions of introns, suggests this is unlikely. Vertebrates such as mouse and *Xenopus* have more than one *caudal* gene homologue, possibly as a result of duplication during evolution. It is quite possible that the *C. helgolandicus caudal* gene may have undergone replication and divergence during evolution. Only further analysis of clone 11-RTPCR and clone 20-RTPCR will provide this answer. Interestingly, the flour beetle *Tribolium* has an unusual *caudal* gene arrangement, similar to that seen in *C. helgolandicus*. In *Tribolium* there is partial duplication of the *caudal* gene, resulting in alternative transcripts with identical 5'-exons, but different 3'-exons, encoding two different homeoboxes (Schulz *et al.*, 1998).

The third putative *C. helgolandicus* developmental gene obtained through targeting time specific expression in eggs is referred to as clone 12-IVPCR. Through comparison of sequence identity it would appear that this gene fragment is a *C. helgolandicus* homologue belonging to the *Antennapedia* class of genes, also referred to as *Hox* genes. Genes within this *Antennapedia* complex are homeotic genes, defined as genes which, when mutated, cause transformations of one part of the body into another. Clone 12-IVPCR has a similar sequence to the chick *HoxB3* gene (and other members of the paralogous subgroup of vertebrate Hox genes), the amphioxus gene *AmphiHox3* and the *Drosophila Sex Combs reduced* gene (*Scr*). Detailed comparison of clone 12-IVPCR with the above mentioned sequences shows that the highest homology is within the homeodomain or the hexapeptide, located just upstream of the homeodomain. Interestingly, *AmphiHox3* has an intron

between the homeodomain and the hexapeptide, this intron location is shared with the mouse *HoxB3* gene (Holland *et al.*, 1992), but not with the chick *HoxB3* gene (Rex and Scotting, 1994). Also of interest is the fact that the mouse and human *HoxB3* sequences predict a stretch of 26 glycine residues (interrupted by two serine residues) between the hexapeptide and the homeodomain. This unusual feature is not present in either *AmphiHox3*, chick *HoxB3*, or the *C. helgolandicus* clone 12-IVPCR sequence.

Clone 12-IVPCR is certainly only a partial sequence, as comparison with the chick *HoxB3* gene of 399 amino acids and *Scr* gene of 413 amino acids confirms. Unfortunately, comparison with homologous genes cannot provide an explanation for the stop codon found 96 amino acids after the homeodomain. There are no obvious indications that this is either an intron or the terminating codon for the gene.

Many of the genes discussed here belong to a 3' paralogous subgroup of vertebrate Hox genes, such as chick *HoxB3* and mouse *HoxB3*. It had been suggested by Holland *et al.* (1992) and Rex and Scotting (1994) that this anteriorly expressed vertebrate subgroup of the *Antennapedia* complex, containing *HoxA3*, *HoxB3*, and *HoxD3* genes, is homologous to the *Drosophila proboscipedia* gene. However, Duboule (1994) claims that this paralogous subgroup is mistakenly referred to as a *proboscipedia* homologue. He suggests that this group may in fact represent a unique vertebrate paralogous subgroup, or a *Drosophila* gene that has been lost during evolution. If this subgroup is unique to vertebrates it is confusing to see such a strong sequence similarity, and indeed to some extent, an expression similarity between the *AmphiHox3* and murine *HoxB3* genes (Holland *et al.*, 1992). Certainly further analysis of clone 12-IVPCR is needed to predict where this *C. helgolandicus* gene fits into the complex organisation of the gene clusters in the *Antennapedia* complex. Interestingly, Averof and Akam (1993) found an *Artemia* gene that, although related to the *Antennapedia*-like HOM/Hox family, was not clearly identifiable as the specific homologue of any known gene. They suggest that this, and

similarly divergent *Antennapedia*-like homeoboxes isolated from insects, may compromise a family of rapidly diverging homeobox genes.

With all the sequence comparisons made it would be fair to suggest that clone 12-IVPCR is certainly a *C. helgolandicus* developmental gene belonging to the *Antennapedia* complex, this alone giving an insight into its possible expression. *Antennapedia* genes are master genes, involved in determination. They control the position-specific differentiation of body parts. It is now accepted that Hox genes are regulators of anteroposterior specification, in animal groups ranging from *Drosophila* to vertebrates (Slack *et al.*, 1997; Holland and Garcia-Fernandez, 1996). The chick *HoxB3* gene was cloned from 2.5-day-old embryos (Rex and Scotting, 1994). The expression of *AmphiHox3* gene in *B. floridae* embryos and larvae ranges from 13 hours (the earliest embryonic stage analysed) to 7 days post fertilisation (Holland *et al.*, 1992). The *Drosophila homeotic* gene *Scr*, to which clone 12-RTPCR shows a high degree of homology, was isolated from 0-22 hour embryos (LeMotte *et al.*, 1989). These patterns of detection in the *Antennapedia* developmental genes, that show considerable homology to the *C. helgolandicus* clone 12-IVPCR sequence, confirm that clone 12-IVPCR is an exciting candidate for further analysis. It seems quite possible that clone 12-IVPCR will be expressed post-fertilisation and throughout early embryogenesis. This *C. helgolandicus* developmental gene is expressed in *C. helgolandicus* eggs. If further analysis of its temporal expression in developing embryos were to prove that the gene is expressed within the first 48 hours of development, the gene could be used to construct suitable probes for the *in situ* analysis of expression within copepod eggs.

All three genes, clone 11-IVPCR, clone 20-IVPCR and clone 12-IVPCR, acquired through targeting temporal-specific analysis, provide exciting opportunities to investigate the effect of external parameters such as food quality, and quantity on developmental gene expression.

CHAPTER 6

Development of a molecular technique to distinguish the identity of *Calanus* species at any developmental stage

CHAPTER 6

Development of a molecular technique to distinguish the identity of *Calanus* species at any developmental stage

6.1 Introduction

6.1.1 Why distinguish between the North Atlantic *Calanus* species?

6.1.2 Why use genetic characters for taxonomic discrimination?

6.1.3 Genetic strategies for species identification

6.2 Methods

6.2.1 Sample collection and preservation

6.2.2 Development of a Restriction Fragment Length Polymorphism (RFLP) technique to distinguish between *Calanus* species

6.2.2.1 Design of amplicon for RFLP analysis

6.2.2.2 Amplification of 16S rDNA fragments from Calanus

6.2.2.3 Cloning and sequencing of 16SAR - 16SB2R Calanus *fragments*

6.2.2.4 RFLP analysis of amplified products

6.2.3 Extension of the technique for use on individual body parts and at any developmental stage

6.3 Results

6.3.1 Sample preservation

6.3.2 Development of a Restriction Fragment Length Polymorphism (RFLP) technique to distinguish between *Calanus* species

6.3.2.1 Design of amplicon for RFLP analysis

6.3.2.2 Amplification of 16S rDNA fragments from Calanus

6.3.2.3 Cloning and sequencing of 16SAR - 16SB2R Calanus *fragments*

6.3.2.4 RFLP analysis of amplified products

6.3.3 Extension of the technique for use on individual body parts and at any developmental stage

6.4 Discussion

6.1 Introduction

6.1.1 Why distinguish between the North Atlantic *Calanus* species?

One of the persistent problems when investigating marine zooplankton communities, especially copepod communities, relates to the correct identification of the different species present. Whilst many adult stages of copepod can be identified to species level, their earlier developmental stages present a persistent identification problem to zooplankton ecologists. For example, the use of *Calanus helgolanidicus* for molecular developmental studies demonstrated a need for an accurate and unambiguous technique to distinguish *C. helgolandicus* from other species, especially *C. finmarchicus*. Despite the relatively high abundance and ecological importance of members of the *Calanus* genus, in both coastal and open oceanic planktonic assemblages, identification between different species is still problematic.

The difficulty in discriminating between the morphologically similar species of *Calanus* prompted the development of a molecular system to identify accurately between the four species *C. helgolandicus*, *C. finmarchicus*, *C. glacialis* and *C. hyperboreus*. It is important to distinguish between these four *Calanus* species because they have different life history strategies, different vertical distributions, opposite temperature affinities and yet they co-occur and compete interspecifically. These four species of *Calanus* were chosen for study, not only due to their co-occurrence, but also because they constitute a major component of zooplankton in the N. Atlantic and N. Sea in terms of biomass and trophic role. The species respond differently to environmental conditions, and to understand how it is necessary to accurately describe their spatial and temporal pattern. To achieve this there is a need to discriminate unambiguously between the species at any developmental stage.

6.1.2 Why use genetic characters for taxonomic discrimination?

Calanus species are reproductively isolated but demonstrate extreme morphological similarity (Fleminger and Hulsemann, 1977; Bucklin *et al.*, 1995), thus presenting a persistent problem in the identification of individuals to species level. Diagnostic morphological characteristics of *Calanus* species are restricted essentially to minor variations in their secondary sex characteristics. For example, the structure of the fifth pair of swimming legs distinguishes adult males of *C. finmarchicus* from *C. helgolandicus*. From a ventral view of the animal, the right leg of the fifth pair of swimming legs for *C. helgolandicus* is shorter than the left leg; the end of the third exopod on the right leg aligning with the second exopod on the left leg. For *C. finmarchicus,* the right leg is also shorter than the left, but the difference in length is less than for *C. helgolandicus*; the third exopod of the right leg aligning half way down the third exopod of the left leg. The fifth pair of swimming legs also distinguishes adult females of *C. finmarchicus* from *C. helgolandicus*. The diagnosis however, is only apparent after detailed examination of the first segment of the leg (coxa). In *C. helgolandicus,* the inner coxa is concave whereas in *C. finmarchicus* the coxa is straight/slightly convex (Figure 6.1).

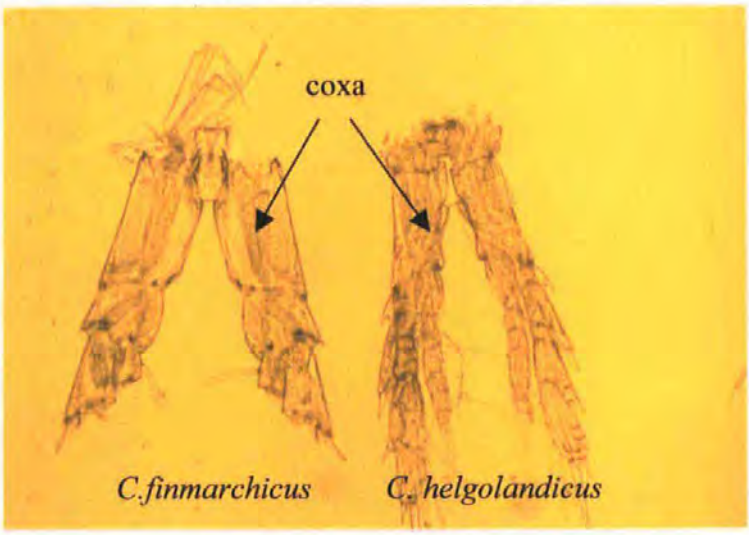


Figure 6.1 Fifth pair of swimming legs of adult female *Calanus helgolandicus and Calanus finmarchicus*. The inner shape of the coxa is used to distinguish between *C. finmarchicus* (straight or convex), and *C. helgolandicus* (concave).

As identification of *Calanus* to species level relies primarily on small variations in the secondary sex characteristics, identification of individuals is feasible only in adult and late copepodite stages. Immature *Calanus* are difficult to assign to a species, and historically, the geographic location of collection has, at least partially, been relied upon as an indicator of species identity. The dangers associated with such an assumption are obvious. When morphological characters are not practical for routine identification to species level, characteristics conserved within their genetic makeup can be used to provide unambiguous taxonomic discrimination at any developmental stage.

6.1.3 Genetic strategies for species identification

Various biochemical and molecular approaches can be applied to resolve problems of species identification and/or intraspecific variation. Such methods rely on the ability to discriminate between variants of different genes or between intergenic regions.

Historically, detection has been directed towards specific gene products (e.g. proteins). The first biochemical strategies used for systematic studies involved enzyme polymorphisms. Manwell *et al.* (1967) used electrophoretic techniques to study protein and enzyme polymorphisms to distinguish between different calanoid copepods. A similar method of isoelectric focusing was used to determine the variability of glucose phosphate isomerase in individual copepods of six species of the genus *Calanus* (Sevigny and McLaren, 1988). Allozymic variation has also been used to assess the population genetic structure and patterns of dispersal of marine zooplankton (Bucklin and Marcus, 1985, Bucklin and Wiebe, 1986, Bucklin *et al.*, 1989, Bucklin, 1991).

Whilst the study of allozymes has been useful for determination of species and population genetic structure, the large amount of material required for protein studies prohibits the use of this technique for routine species identification. Recent molecular approaches, such as DNA sequence variation, can provide an alternative to protein

methods for systematic studies. For example, the sequences of mitochondrial DNA regions have been used for *Calanus* population genetic studies (Bucklin *et al.*, 1992; Bucklin and Lajeunesse, 1994; Bucklin *et al.*, 1995; Bucklin *et al.*, 1996; Bucklin and Kocher, 1996). Sequencing a region of DNA from individual animals however, is also inappropriate for use as a routine technique for species identification as it is too time consuming and expensive.

Other diagnostic molecular techniques used previously for species identification and population studies include methods based on random genomic analysis, or DNA fingerprinting. Such methods analyse simultaneously multiple loci across the genome, usually by PCR using suites of arbitrarily-designed oligonucleotide primers. One genomic fingerprinting approach is Random Amplified Polymorphic DNA (RAPD; Welsh and McClelland, 1990, Williams *et al.*, 1990). RAPD techniques, involving amplification of genomic DNA by arbitrary primers, have been used extensively, for example, to reveal population genetic structure among crustaceans (Martinez *et al.*, 1997; Todd *et al.*, 1997).

An alternative molecular approach to systematic studies is fragment length polymorphism (FLP) analysis. Such techniques involve the analysis of different length fragments of DNA particular to different populations or species. The DNA fragments are produced by various means and from either genomic DNA or particular regions of DNA. One such RFL technique is Amplified Fragment Length Polymorphism (AFLP) analysis. AFLP can take many forms, being used on single or multiple loci. However, for species identification, the AFLP technique based on random genomic analysis is often used (Vos *et al.*, 1995). This AFLP analysis involves amplifying selectively a subset of genomic restriction fragments using specially-designed PCR primers. For example, the AFLP technique has been used to evaluate differentiation of *Artemia* species (Triantaphyllidis *et al.*, 1997). Although the AFLP technique described by Vos *et al.* (1995) has been considerably simplified recently (Suazo and Hall, 1999), both the described AFLP and

RAPD random genomic analysis techniques require prior purification of the DNA to obtain a reproducible result. Such a purification step would take time and preclude the analysis of early developmental stages, such as single eggs and nauplii, due to limiting amounts of starting material. RAPD and AFLP techniques also involve a number of methodology steps so increasing the difficulty and cost.

A second variation of FLP analysis is Restriction Fragment Length Polymorphism (RFLP) analysis. RFLP involves digesting DNA with chosen endonuclease restriction enzymes. The enzymes are selected to digest differentially the chosen region of DNA, resulting in distinct profiles of differing length DNA fragments for each species under study. A variation of RFLP was the chosen method used in this study for the identification of *Calanus* species because of its relatively low cost, simplicity and its suitability for use on small amounts of starting material when preceded by PCR amplification.

The region of DNA chosen for RFLP analysis is that of the mitochondrial 16S ribosomal RNA gene (16S rDNA). The 16S rRNAs of diverse organisms, although varying in their nucleotide sequences, still contain regions that are conserved (Woese, 1987). These regions have been used for the design of universal 16S primers (Palumbi and Benzie, 1991), allowing the amplification of a region of the 16S rRNA gene from many different organisms. Amplification and analysis of the 16S rRNA gene has been used to determine phylogenetic relationships in a wide range of organisms (Lane *et al.*, 1985; DeLong *et al.*, 1989; Rehnstam *et al.*, 1989; Ward *et al.*, 1990; Britschgi and Giovannoni, 1991). The nucleotide sequence divergences in the gene portion of 16S rDNA amplified from *Calanus* species, with the universal primers 16SAR and 16SBR, have been used for intraspecific and interspecific studies. These sequence divergences among *C. finmarchicus*, *C. marshallae* and *C. pacificus* range from 12 – 19% (Bucklin *et al.*, 1992), and between *C. glacialis*, *C. marshallae*, *C. sinicus* and *C. hyperboreous* between 7.3% and 23.9% (Bucklin *et al.*, 1995). However, between geographically distant conspecific

populations the variation is normally less than 1% (Bucklin *et al.*, 1992) and has never been found to be higher than 2% (Bucklin *et al.*, 1995; Bucklin and Wiebe, 1998). In *Calanus* species, the intraspecific variation of mitochondrial 16S rDNA is small compared with other species of crustaceans, despite the enormous population sizes, short generation times, and broad geographic distributions of *Calanus* (Bucklin *et al.*, 1995; Bucklin and Wiebe, 1998). Although the maternal inheritance of 16S rDNA is advantageous in elucidating population structure within species, because it allows discrimination and identification of maternal lineages, the use of 16S rDNA for population genetics is limited as it has shown too little genetic divergence on all but a macroscale (Bucklin *et al.*, 1996).

The relatively low intraspecific variation within the amplified portion of 16S rDNA however, in addition to the sufficient base sequence divergence exhibited by *Calanus* species within this region, confirms the suitability of the 16S rRNA gene for *Calanus* species identification. The technique developed involves PCR amplification of a region of the mitochodrial 16S rRNA gene without prior purification of the DNA (Bucklin *et al.*, 1995). The amplification of this gene fragment is followed by RFLP analysis. In this study, the work by Bucklin *et al.* (1992; 1995), in which the DNA sequence of 16S rRNA genes has been determined and used to compare phylogenetic relationships between *Calanus* species, has been extended. This system has been further developed to produce a method whereby simple molecular techniques can be used to determine unambiguously the species of adults, juveniles and eggs of the dominant *Calanus* species found in the N. Atlantic. Additionally, this newly developed technique identifies *Calanus* to species based on small body parts, enabling identification of species from gut content analysis of *Calanus* predators.

6.2 Methods

6.2.1 Sample collection and preservation

Individuals of *C. helgolandicus*, *C. finmarchicus*, *C. glacialis* and *C. hyperboreus*, collected by net tows (WP2 net, mesh 200μm, UNESCO 1968), were preserved in absolute ethanol, at room temperature, with approximately 10 adults per 10mL ethanol. *Calanus helgolandicus* were collected (April 1996) from a coastal station approximately 10 km off Plymouth, western English Channel (50°15'N, 4°13'W); *C. finmarchicus* were collected (April/May 1996) from Ganavan Bay, Oban, Western Scotland (56°27'N, 5°27'W); and *C. glacialis* and *C. hyperboreus*, supplied by Dr. H. J. Hirche, were collected (June 1997) from the Norwegian Sea (77°01'N, 33°50'E).

6.2.2 Development of a Restriction Fragment Length Polymorphism (RFLP) technique to distinguish between *Calanus* species

6.2.2.1 Design of amplicon for RFLP analysis

Fragments of 16S rDNA have previously been amplified from different *Calanus* species (Bucklin *et al.*, 1992, Bucklin 1995) using universal primers to the 16S rRNA gene (Palumbi and Benzie, 1991). However, the published sequence data for these amplified 16S rDNA regions, bordered by the universal primers 16SAR and 16SBR, is incomplete. Therefore, to maximise the potential for RFLP development such that fragments could be resolved suitably by agarose gel electrophoresis, the complete amplified sequence was obtained. From the four *Calanus* species under study, 16S rDNA was amplified with 16SAR and 16SBR primers following the methods detailed elsewhere (Section 6.2.2.2). For each of the four *Calanus* species, the amplified 16S rDNA fragments were cloned into pBluescript SK^- and sequencing was achieved following the chain termination method (Section 2.3). Sequencing was performed in both directions on at least 2 clones for each fragment.

Using the obtained sequence information, a new reverse amplification primer was designed (16SB2R 5'-ATTCAACATCGAGGTCACAAAC-3'). 16SB2R was designed to produce a suitable amplicon for RFLP. The Tm of 16SB2R was closer to that of the forward 16SAR primer than the original reverse 16SBR primer, therefore allowing a more effective amplification. The Genetics Computer Group (GCG) Sequence analysis Software Package (Devereux *et al.*, 1984) was used to restriction map the 16SAR-16SB2R sequences for *C. helgolandicus, C. finmarchicus, C. glacialis and C. hyperboreus.* From the restriction maps a combination of restriction enzymes, compatible in a single reaction buffer, were chosen. The enzymes were selected to digest differentially the amplified 16S rDNA fragments, producing a unique discernible pattern for each *Calanus* species.

6.2.2.2 Amplification of 16S rDNA fragments from Calanus

Amplifications of the 16S rRNA gene were performed on single, preserved animals without prior purification of the DNA (Bucklin *et al.* 1995). Individual animals were removed from ethanol, and rehydrated in 0.5 mL of distilled water in a microcentrifuge tube for 6-12 h at room temperature. After rehydration, the water was removed and replaced by 34 µL MilliQ water and 5 µL of 10x Dynazyme buffer. The animals were homogenised using a pellet pestle homogeniser and incubated at 4°C overnight. After incubation, the homogenate was transferred to a 0.7 mL tube and the remaining reaction components added:-

2 mM dNTPs	- 5 µL
100 ng. μL^{-1} 16SAR	- 2.5 µL
100 ng. μL^{-1} 16SB2R	- 2.5 µL
Dynazyme (2U μL^{-1})	- 1 µL

The amplification primers were:-

16SAR - 5'CGCCTGTTTAACAAAAACAT-3' (Palumbi and Benzie, 1991)

16SB2R - 5'-ATTCAACATCGAGGTCACAAAC-3' (custom designed from *Calanus* sequences).

Amplification was carried out in an Autogene thermocycler (Grant). The cycling parameters included an initial denaturation step of 96°C (5 min) followed by 40 cycles of 45°C (2 min), 72°C (1 min), and 96°C (1 min). A final annealing phase at 45°C (2 min) was followed by an extension phase at 72°C (3 min) and storage at 4°C until use. 10μL aliquots of the amplification reactions were analysed by agarose gel electrophoresis (1.5%) to check the efficiency of amplification.

6.2.2.3 Cloning and sequencing of 16SAR - 16SB2R Calanus *fragments*

To confirm the identity of the fragments amplified from the four *Calanus* species with the 16SAR and 16SB2R primers, fragments were cloned into pBluescript SK⁻, and sequenced following the chain termination method (Section 2.3). Sequencing was performed in both directions on a minimum of 2 clones for each fragment.

6.2.2.4 RFLP analysis of amplified products

Restriction digests were performed on a 15 μL aliquot from each amplification, by the addition of 0.5 μL of 5M NaCl, 2 μL bovine serum albumin (1 mg mL^{-1}) and 5 units of each selected restriction enzyme (*Vsp* I, *Dde* I). Incubations were performed at 37°C for 1 h. The digestion products were separated by electrophoresis through a 2% Metaphor agarose gel, pre-chilled for 30 min at 4°C to improve resolution. The gels were observed and photographed by UV transillumination.

6.2.3 Extension of the technique for use on individual body parts and at any developmental stage

16S rDNA fragments were amplified, using the 16SAR and 16SB2R primers, from all life stages of a species, following the amplification procedure described in Section 6.2.2.2. The technique was applied to individual eggs, all nauplii stages and all copepodite stages. The amplification was also performed on small body parts including the antennule,

leg, thorax and urosome. In many cases, because of the small amount of material being used, the tissue was rehydrated in approximately 200 µL of MilliQ water, rather than the described 500 µL. The smaller quantity of water allowed easier retrieval of the tissue following rehydration. The amplified fragments were restricted following the protocol described in Section 6.2.2.4. The digestion products were separated by electrophoresis through a 2% Metaphor agarose gel and the resulting restriction profiles viewed by UV transillumination.

6.3 Results

6.3.1 Sample preservation

Samples stored in absolute ethanol were adequately preserved, and proved suitable for DNA amplification after rehydration. As yet (3 years) there appears to be no limit to the length of time for which animals can be preserved. In fact, it was necessary to preserve the animals in ethanol for a minimum of 12 h, presumably helping to disrupt the structural components prior to homogenisation.

6.3.2 Development of a Restriction Fragment Length Polymorphism (RFLP) technique to distinguish between *Calanus* species

6.3.2.1 Design of amplicon for RFLP analysis

An amplicon was successfully designed for RFLP analysis. Cloning and sequencing of 16SAR and 16SBR fragments allowed the design of a suitable new reverse amplification primer (16SB2R). The primer was designed to a region of high conservation, allowing amplification from all four *Calanus* species, and with a Tm closer to 16SAR than the original 16SBR primer, allowing a more effective amplification.

The sequenced rDNA regions were mapped for the presence of suitable RFLPs. The chosen restriction enzymes *Dde*I and *Vsp*I were optimally active in the amplification

buffer with the addition of 0.5 µL 5M NaCl. The positions of the restriction sites for *Dde*I and *Vsp*I on the amplified 16S rDNA fragment for each of the four *Calanus* species are shown in Figure 6.2.

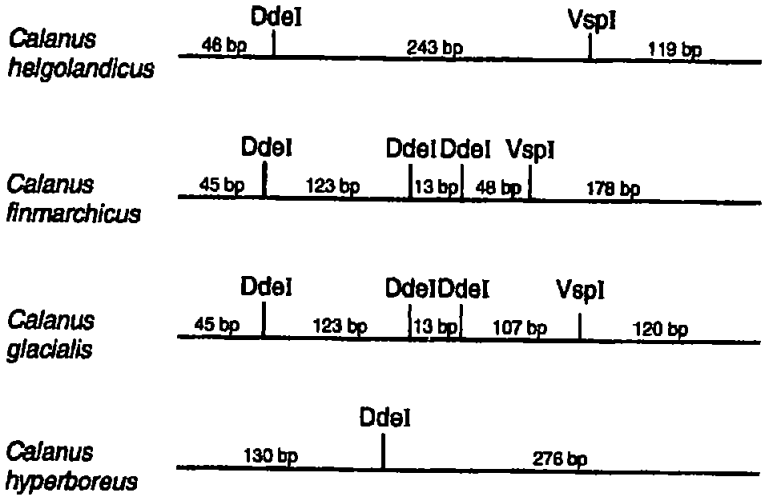


Figure 6.2 Restriction map of the amplified region of mitochondrial 16S rDNA, with 16SAR and 16SB2R primers, for *Calanus helgolandicus*, *Calanus finmarchicus*, *Calanus glacialis*, and *Calanus hyperboreus*. Both *Dde*I and *Vsp*I restriction sites are indicated. The resulting fragment sites are shown in base pairs (bp).

6.3.2.2 Amplification of 16S rDNA fragments from Calanus

The amplification primers 16SAR and 16SB2R were used routinely to amplify a 411 bp region of 16S rDNA from single preserved *Calanus* individuals (Figure 6.3).

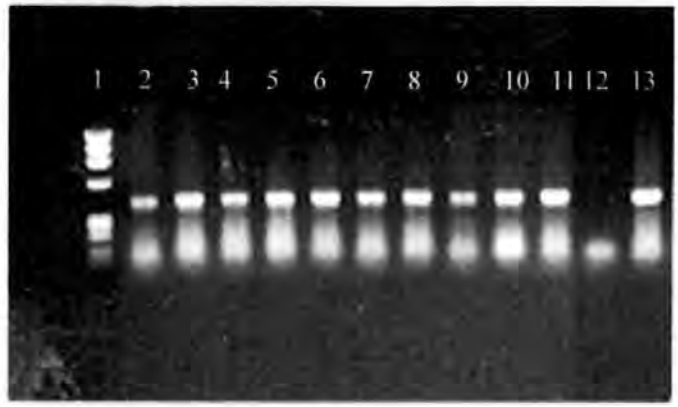


Figure 6.3 1.5% agarose gel showing amplification of 16S rDNA, with 16SAR and 16SB2R primers, from whole adults. Lane 1 = 200 ng PhiX174/*Hae*III molecular weight marker (size of fragments in bp = 1353, 1078, 872, 603, 310, 281 and 271, 234, 194, 118, 72); lanes 2 – 11 = 10 µL aliquots of the amplification reactions from individual *Calanus*; lane 12 = -ve control; lane 13 = +ve control (200 ng *Calanus helgolandicus* purified DNA).

6.3.2.3 Cloning and sequencing of 16SAR - 16SB2R Calanus *fragments*

Nucleotide-base sequences for the 411 bp amplified region of mitochondrial 16S rDNA were determined for all four *Calanus* species (Figure 6.4).

```
  chel16S  cgcctgttta acaaaaacat cgtaaaatag atttatattg cctgctcagt
  cfin16S  cgcctgttta acaaaaacat cgtaaataat atttataata cctgctcagt
 cglac16S  cgcctgttta acaaaaacat cgtaaataat atttataata cctgctcagt
chyper16S  cgcctgttta acaaaaacat .gtaaattag atttataatg cctgctcaat

  chel16S  g.aatattta aacagccgca ttagtgttaa ggtagcatag taattagttt
  cfin16S  ..aatattta aacagccgcg ttagtgttaa ggtagcatag taattagttt
 cglac16S  g.aatattta aacagccgcg ttagtgttaa ggtagcatag taattagttt
chyper16S  gaaatattta aatagccgcg ttagtgttaa ggtagcatag taattagttt

  chel16S  tttaattgga aaatggaatg aatggcccca ctaaagcata gtatttatac
  cfin16S  cttaattggg aaataggatg aatggtttca ctaaaatata gtttttatcc
 cglac16S  cttaattggg aaatggaatg aatggttt.a ctaaaatatg atatttattc
chyper16S  tttaattgga aaatggaatg aatggtttca ctaagatatg gtatttatgc

  chel16S  taaaaaatga aattttaatt taagtgaaaa tacttaaatg atatatttag
  cfin16S  tcatttgcga aattttaatc taagtgaaaa tacttagcag ttgtactagg
 cglac16S  taatttgcga aattttaatc tgagtgaaaa tactcagaag atatatttag
chyper16S  taataaatga aattttaatt taagtg.aaa tacttaaaag atcttttaag

  chel16S  acgagaagac cctatgaagc tggtagacca taagtgtaat tatttcatag
  cfin16S  acgagaagac cctatgaagc tggcaaacta ttaat.acat attcctatta
 cglac16S  acgagaaga. cctatgaagc tggtagactt ccaatgtaat tatacgatag
chyper16S  acgagaagac cctatgaagc t.atagacta taaatataat tattataaat

  chel16S  tag.caggtc tatttttgg ggtaaaattt aataattata ttaatacaga
  cfin16S  tttattagtt tatttttgg ggtaaaattt aataatacta ttaacacaat
 cglac16S  ttcatgagtt tatttttgg ggtaaaattt aataatagta ttaatattgg
chyper16S  ttt.taagtt tatttttgg ggtaaaattt aataatttta tttaaataag

  chel16S  tttgttcaaa cttatccttc tggaattatg aataagctcc tctagggata
  cfin16S  tgtactaaat tacatccttt aggaattatg aagaagctcc tctagggata
 cglac16S  cttactaaat aatatcctct tggaattatg aaaaagctcc tctagggata
chyper16S  cttattttaa ttgatccttt aggaattatg aaaaagctcc tctagggata

  chel16S  ac.agcatta tgcttaaaag agttcttatc agaataagcg tttgtgacct
  cfin16S  ac.agcatta tgcttaaaag agttcttatc agaataagcg tttgtgacct
 cglac16S  acaagcatta tgcttaaaag agttcttatc agaataagcg tttgtgacct
chyper16S  ac.agcatta tacctataag agttcttatc agaataggtg tttgtgacct

  chel16S  cgatgttgaa t
  cfin16S  cgatgttgaa t
 cglac16S  cgatgttgaa t
chyper16S  cgatgttgaa t
```

Figure 6.4 Sequence data for a 411 bp region, bordered by 16SAR and 16SB2R primers, of the mitochondrial 16S rRNA gene. cfin = *Calanus finmarchicus*, cglac = *Calanus glacialis*, chel = *Calanus helgolandicus*, and chyper = *Calanus hyperboreous*. (. = alignment gap; n = any base; **ctnag** = *Dde*I restriction site; **attaat** = *Vsp*I restriction site; primers are denoted by italics).

6.3.2.4 RFLP analysis of amplified products

The amplification of 16S rDNA and subsequent RFLP analysis produced a characteristic pattern for each *Calanus* species, providing an unambiguous statement of identity (Figure 6.5).

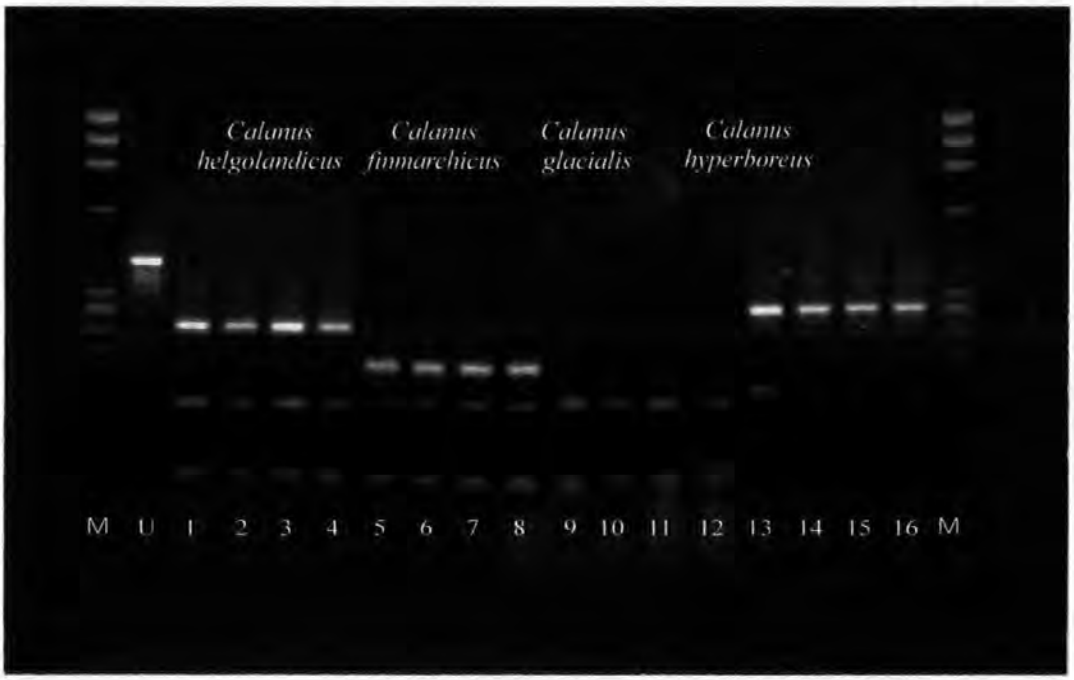


Figure 6.5 Amplification of 16S rDNA from whole adult animals and subsequent RFLP analysis with restriction endonucleases *Dde*I and *Vsp*I. Lanes 1 - 4 = *Calanus helgolandicus*; Lanes 5 - 8 = *Calanus finmarchicus*; Lanes 9 - 12 = *Calanus glacialis*; Lanes 13 - 16 = *Calanus hyperboreus*. M = phiX174/*Hae*III DNA size marker; U = amplified uncut *Calanus helgolandicus* DNA.

6.3.3 Extension of the technique for use on individual body parts and at any developmental stage

16S rDNA fragments were amplified successfully, using the 16SAR and 16SB2R primers, from eggs, nauplii and copepodites. Subsequent RFLP analysis gave unambiguous restriction profiles for each species. The versatility of this molecular method, allowing all life stages of a species to be identified, is demonstrated in Figure 6.6.

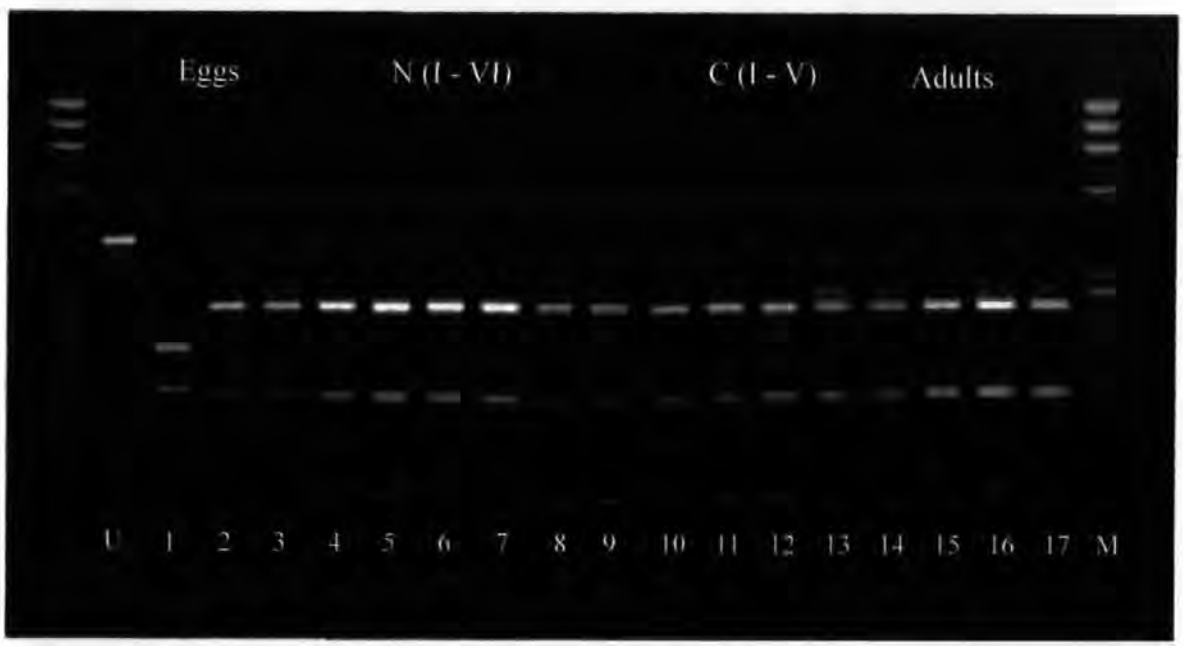


Figure 6.6 Amplification of 16S rDNA from the major developmental stages of *Calanus helgolandicus* and subsequent RFLP analysis with *Dde*I and *Vsp*I. The restriction pattern produced by *Calanus finmarchicus* is shown for comparison (Lane 1). Lanes 2 - 4 = individual eggs; Lanes 5 - 10 = nauplius stages I-VI respectively; Lanes 11 - 15 = copepodite stages I-V respectively; Lanes 16 & 17 = adult animals. M = phiX174/*Hae*III DNA size marker; U = amplified uncut *Calanus helgolandicus* DNA.

This method of species identification is also valid for small body parts, including the antennule, leg, thorax and urosome (Figure 6.7).

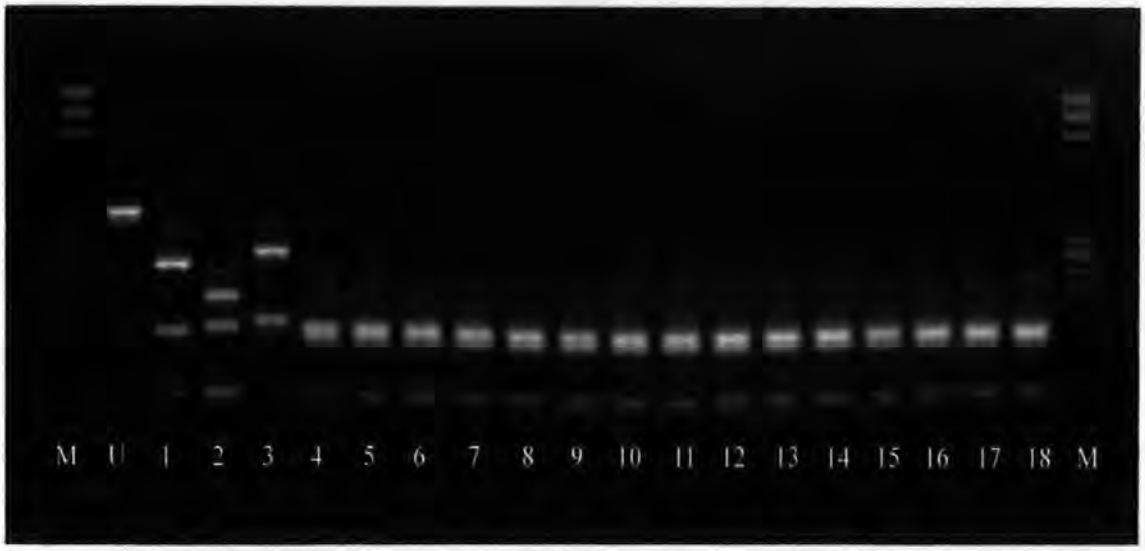


Figure 6.7 Amplification of 16S rDNA from body parts of *C. glacialis* and subsequent RFLP analysis using *Dde*I and *Vsp*I. The restriction pattern produced by *C. helgolandicus* (Lane 1), *C. finmarchicus* (Lane 2) and *C. hyperboreus* (Lane 3) are shown for comparison. Lanes 4 - 6 = individual antennules; Lanes 7 - 9 = individual legs; Lanes 10 - 12 = individual urosomes; Lanes 13 - 15 = individual thoraxes; Lanes 16 - 18 = individual eggs. M = phiX174/*Hae*III DNA size marker; U = amplified uncut *C. helgolandicus* DNA.

6.4 Discussion

A simple molecular technique has been developed, involving amplification of 16S rDNA and subsequent RFLP analysis, which is capable of distinguishing between *Calanus helgolandicus*, *Calanus finmarchicus*, *Calanus glacialis and Calanus hyperboreus* at any life stage. The method is preferable to the RAPD technique and multi loci AFLP techniques as it does not require prior purification of the DNA in order to obtain an accurate and reproducible result. Such a purification step would preclude the analysis of

single eggs and nauplii due to limiting amounts of starting material, and also extend the overall time required for analysis.

Existing 16S rDNA amplification systems (Palumbi and Benzie, 1991; Bucklin *et al.*, 1992) have been modified to provide a cost effective PCR product suitable for subsequent RFLP analysis. The newly designed 16SB2R primer has a predicted Tm closer to that of the forward 16SAR primer, allowing the annealing temperature used in the PCR to be increased to 45°C. The higher Tm increases the specificity of the amplification product, which in turn produces a more distinct restriction profile. The reaction volume has been decreased by 50% to reduce the cost of reactants, and the extension phase of thermocycling reduced to save time.

The restriction enzymes were selected not only for the production of a different restriction profile for each species, but also for their compatibility of reaction buffer, such that simultaneous double-digestion could be performed in a single reaction. Additionally, there was no requirement for purification of the amplified product or buffer exchange prior to restriction digestion. Simply by increasing the salt concentration of the PCR buffer by the addition of NaCl, a suitable buffer for activity of the two restriction enzymes is created. This system, involving the amplification of rDNA sequence from single animals and the subsequent RFLP analysis, provides an accurate and reliable diagnostic technique with which to identify *Calanus* species.

The sequence data obtained for 16S rDNA fragments amplified from the four *Calanus* species differed slightly from previously published sequence data for the same region of 16S rDNA for the same species (Bucklin *et al.*, 1995). For *C. helgolandicus,* variation was low compared with *C. helgolandicus* collected previously from the English Channel (1%) and from the Azores (1.7%). Present sequence data for *C. finmarchicus* also varied only slightly (0.3%) from *C. finmarchicus* collected from the Gulf of Maine and from the Gulf of St. Lawrence (2.3%). *Calanus hyperboreus* showed 2.3% variation from

C. hyperboreus collected in the Gulf of St. Lawrence. *Calanus glacialis* showed greater variation (9.9%) compared with the same species collected in the Gulf of St. Lawrence. The intraspecific variation is small in the case of *C. helgolandicus*, *C. finmarchicus* and *C. hyperboreus*, but significant for *C. glacialis*. The latter difference may reflect the site of collection, with original sequence data being obtained for *C. glacialis* from the Gulf of St. Lawrence and present sequence data being obtained from *C. glacialis* collected from the Norwegian Sea. Although this sequence variation is not as significant at that recorded between species (Bucklin and Lajeunesse, 1994), it does pose some interesting questions about the phylogenetic relationships of geographically diverse *Calanus* species.

There are many practical applications for this simple molecular technique. It has been well demonstrated that environmental factors such as temperature, oceanic currents, turbulence, wind stress and nutrients have considerable influences on zooplankton population composition (Colebrook, 1982, 1985; Dickson, 1988; Fromentin and Planque, 1996), and that different combinations of these species co-occur (Fleminger and Hulsemann, 1977; Frost, 1974; Williams and Conway, 1980). An unambiguous method of distinguishing between *Calanus* species at any life stage will increase understanding of the effect of these environmental factors on zooplankton assemblages, thus enabling a better understanding of their distribution, vertical migration and exploitation of habitat. This in turn will permit further investigations into how environmental changes affect individual zooplankton species having different physiological and biological properties.

Copepods, particularly their eggs and nauplii, play an important role in marine food webs by providing an important food source for many animals including fish larvae, small species of fish and euphausiids (Runge, 1988; Runge and Lafontaine, 1996). Such conclusions are based on investigation of gut contents. *Calanus* eggs can remain undigested during gut passage (Conway *et al.*, 1994) and can, therefore, be identified easily using PCR and RFLP analysis. For euphasiids however, food is well masticated and

only small body parts of *Calanus* are found in the gut. In such cases, this new molecular method enabling identification of species from fragmentary material will be particularly valuable.

In conclusion, the developed method for the identification of North Atlantic *Calanus* species is reliable and relatively simple and inexpensive. It has the benefit that analysis is performed on animals preserved in ethanol, making it suitable for use on *Calanus* samples collected at sea.

CHAPTER 7

Application of the molecular technique, developed to distinguish the identity of *Calanus* species, to a range of North Atlantic samples

CHAPTER 7

Application of the molecular technique, developed to distinguish the identity of *Calanus* species, to a range of North Atlantic samples

7.1 Introduction

7.1.1 Molecular system developed for the identification of *Calanus* individuals in the North Atlantic

7.1.2 Application of the technique

7.2 Methods

7.2.1 Collection and preservation of samples

7.2.1.1 Lurefjorden samples

7.2.1.2 Mesocosm samples

7.2.1.3 North Atlantic samples

7.2.2 Molecular identification of *Calanus* individuals

7.2.2.1 Modified DNA amplification technique

7.2.2.2 Restriction Fragment Length Polymorphism analysis

7.2.3 Comparison of morphological and molecular identification of *Calanus* species from Lurefjorden

7.2.4 Molecular identification of *Calanus* individuals in a mesocosm study

7.2.5 Molecular identification of *Calanus* individuals from a wide geographical range in the North Atlantic

7.3 Results

7.3.1 Preservation of samples

7.3.2 Molecular identification of *Calanus* individuals

7.3.3 Comparison of morphological and molecular identification of *Calanus* species from Lurefjorden

7.3.4 Molecular identification of *Calanus* species in a mesocosm study

7.3.5 Distribution of *Calanus* species in the North Atlantic determined by the molecular technique

7.4 Discussion

7.1 Introduction

7.1.1 Molecular system developed for the identification of *Calanus* individuals in the North Atlantic

A relatively simple method using molecular techniques to identify the four species of *Calanus* common to the North Atlantic has been developed (Chapter 6; Lindeque *et al.*, 1999). This system involves the PCR amplification of a region of the mitochondrial 16S rRNA gene, followed by Restriction Fragment Length Polymorphism (RFLP) analysis of the amplified product. Using this system, it is possible to identify to species not only whole animals at any developmental stage from egg to adult, but also individual body parts.

The aim of this chapter is to test comprehensively the accuracy and reproducibility of this diagnostic technique on a number of *Calanus* samples taken from a wide geographical range within the North Atlantic. The 16S rDNA primers (16SAR and 16SB2R), used for the amplification of a portion of this gene from *Calanus*, are complimentary to conserved regions of the 16S gene, ensuring amplification from all four of the *Calanus* species studied. The restriction sites of the chosen enzymes *Dde*I and *Vsp*I are obviously regions of variable sequence between the *Calanus* species under study and are therefore, suitably diagnostic. From detailed comparisons of *Calanus* sequence data from geographically diverse conspecific populations these restriction sites show no intraspecific variation. However, a true test of the intraspecific conservation of the restriction sites, and therefore the developed method, is to apply the technique to *Calanus* samples collected from a wide geographical range.

7.1.2 Application of the technique

The need for an unambiguous method to discriminate between *Calanus* species was illustrated clearly in the study of *Calanus* populations in Lurefjorden (western coast of

Norway). It is suspected that in Lurefjorden, the populations of *Calanus* are not composed of a single species. However, adult *Calanus* in Lurefjorden are of a similar size and therefore harder to separate into species than in other systems (Pers. Comm., Ketil Eiane, 1999). *Calanus* collected from the fjord were attributed to species level, using the morphological characteristics of the curvature of the coxa of the fifth pair of swimming legs. The identification was therefore, restricted to CV stages and adult females. For a non-taxonomist this method is both time consuming and ambiguous. Indeed, the reseacher responsible for the identification was doubtful about the precision level of this approach. It was therefore decided to test the morphological identification of the *Calanus* species, against molecular identification in a blind test of the samples.

The molecular identification technique was also applied to adults and very early developmental stages of *Calanus* in a mesocom study (TASC mesocosm study, Norway, spring 1998). The aim of the mesocosm experiment was to compare the growth and mortality rates of two parallel cohorts of *C. finmarchicus* originating from 60°N (Bergen) and 69°N (Tromsø). By transporting the southern cohort north from its natural habitat, and comparing the development time and stage shifts to the cohort naturally living in the north, it was hoped to determine whether the *C. finmarchicus* life cycle is determined by the environment or if it is strictly genetically programmed. Samples of animals from the mesocosm study were subjected to molecular identification to confirm their initial identification based on morphological traits.

It was of particular interest to obtain *Calanus* samples from areas where species are known to co-occur. Samples were acquired from a wide range of geographic locations, ranging from as far south as the English Channel (50°N) and as far north as Tromsø (69°N). Although the times and regimes of sampling differed significantly, preventing any firm conclusions on *Calanus* distribution to be drawn, it was still of considerable interest to

make a comparison of the distribution of *Calanus* determined by molecular techniques, with traditional distributions that have been established morphologically.

7.2 Methods

7.2.1 Collection and preservation of samples

To obtain *Calanus* samples from a wide geographical area, colleagues in the TASC (Trans Atlantic Study of *Calanus*) community were approached. By asking researchers to preserve aliquots of animals in ethanol, from any fieldwork or cruises, a number of samples were acquired to test the method. In return, molecular identification of the *Calanus* species collected was provided.

7.2.1.1 Lurefjorden samples

Calanus samples were collected from the fjord by vertical net hauls (200 μm WP-2 plankton nets) in both February and April at two depths (0-50 m; 300-400 m), corresponding to a suspected bimodial depth distribution. The animals were stored individually in microcentrifuge tubes containing 1.5 mL of ethanol.

7.2.1.2 Mesocosm samples

Representative samples of animals throughout the mesocosm study were collected by discrete depth water bottle casts. The animals were preserved in ethanol at an approximate concentration of 1 animal per 1 mL of ethanol.

7.2.1.3 North Atlantic samples

Calanus were collected from a wide range of geographical locations in the North Atlantic with 200 μm WP-2 plankton nets. The source, location and method of collection for all of the *Calanus* samples obtained are presented in Table 7.1.

Location	Method of collection	Depth of collection	Source
L4, English Channel	Net tow	50 – 0 m	P. Lindeque
North Sea	Vertical haul	70 – 0 m	S. R. Gonzalez S. S. Oosterhuis
Lurefjorden	Vertical haul	0-50 m 300 – 400 m	E. Bagoien K. Eiane
Faroe shelf	Vertical haul	50 – 0 m	E. Gaard
Offshore areas North and South of Faroes	Vertical Haul	50 – 0 m	E. Gaard
Oban, East Scotland	Vertical Haul	50 – 0 m	P. Lindeque
Weathership M	Vertical haul	100 – 0 m	X. Irigoen
Saltenfjord	Vertical Haul	380 – 0 m	K. Olsen
Bergen and Tromsø	Water bottle casts	Mesocosm	B. Hansen

Table 7.1 Source, location, method and depth of collection of samples to which the molecular method of *Calanus* identification was applied.

The majority of samples were stored in absolute ethanol with approximately 1 animal per 1 mL of ethanol. In some instances (e.g. samples collected around the Faroes shelf) animals were stored at a significantly higher concentration (approximately 100 individuals per 1 mL of ethanol).

7.2.2 Molecular identification of *Calanus* individuals

7.2.2.1 Modified DNA amplification technique

The technique to amplify *Calanus* 16S rDNA with 16SAR and 16SB2R primers (Chapter 6) was modified to reduce both cost and time of analysis. Modifications include the use of *Taq* DNA polymerase in preference to Dynazyme, and a reduction in the amount of restriction enzyme used for each reaction. The final technique is described below:-

Animals were rehydated in 0.5 mL of MilliQ water in a microcentrifuge tube for 6 – 12 h at room temperature. After rehydration, the water was removed and replaced with 34 µL MilliQ water and 5 µL 10 x *Taq* DNA polymerase buffer (Promega UK Ltd.). Copepods were homogenised using a pellet pestle homogeniser and incubated at 4°C

overnight. After incubation, the homogenate was transferred to a 0.7 mL tube and the remaining reaction components were added [5 μL 2 mM dNTPs, 2.5 μL each of primers 16SAR and 16SB2R (100 ng μL^{-1}), and 2 U of *Taq* DNA polymerase (Promega UK Ltd.)]. Amplification was performed in a thermal cycler (PTC-100™, MJ Research, Inc). The cycling parameters included an initial denaturation step of 94°C (5 min) followed by 40 cycles of 45°C (2 min), 72°C (1 min), and 94°C (1 min). A final annealing phase at 45°C (2 min) was followed by an extension phase at 72°C (3 min) and storage at 4°C until use. 10 μL aliquots of the amplification reactions were analysed by agarose gel electrophoresis (1.5%) to check amplification efficiency.

7.2.2.2 Restriction Fragment Length Polymorphism analysis

Restriction digests were performed on a 15 μL aliquot from each amplification by the addition of 0.5 μL 5 M NaCl, 2 μL bovine serum albumin (1 mg mL^{-1}) and 2.5 U of each restriction enzyme (*Dde*I and *Vsp*I). Incubations were performed at 37° C for 1 h. The digestion products were separated by electrophoresis through a 2 % Metaphor agarose gel, pre-chilled for 30 min at 4°C to improve resolution. The gels were observed and photographed by UV transillumination.

7.2.3 Comparison of morphological and molecular identification of *Calanus* species from Lurefjorden

Using the morphological characteristics of the coxa on the fifth pair of swimming legs, Espen Bagoien (Norway) identified 154 *Calanus* collected from Lurefjorden. The ethanol-preserved animals were then individually subjected to molecular identification by the technique detailed in Section 7.2.2, with no prior knowledge of the morphological results.

7.2.4 Molecular identification of *Calanus* individuals in a mesocosm study

Females for the mesocosm study in Tromsø were collected from Raunefjorden in Bergen (March 1998) and Grøtsund in Tromsø (April 1998) by B. Hansen. A representative aliquot of 30 females from each of the two cohorts was identified using the molecular techniques detailed in Section 7.2.2. A further 30 individuals were identified from each cohort at the end of the mesocosm study (June, 1998). At a later date, 20 nauplii from each cohort, taken approximately 1 week into the experiment, were also identified using the molecular technique described (Section 7.2.2).

7.2.5 Molecular identification of *Calanus* individuals from a wide geographical range in the North Atlantic

The geographical areas of collection are shown in Table 7.1. For each of the samples obtained, between 20 and 30 individual animals (adults and late copepodite stages) were picked randomly from storage in ethanol. Each individual was identified using the technique described in Section 7.2.2.

7.3 Results

7.3.1 Preservation of samples

Animals stored at a concentration of 1 individual per 1 mL absolute ethanol were preserved successfully, and proved suitable for DNA amplification. Many samples of *Calanus* stored at a higher concentration (approximately 100 animals per 1 mL ethanol) appeared contaminated with copepod debris, possibly because of physical degradation during transportation. However, intact animals within these samples, following thorough washing in MilliQ water to remove any contaminating fragments, were still suitable for amplification.

7.3.2 Molecular identification of *Calanus* individuals

Over 99% of animals analysed produced a PCR product of the correct size when amplified with the 16SAR and 16SB2R primers. Less than 1% of individual *Calanus* failed to produce any amplification product. All amplified products were digested successfully with *Dde*I and *Vsp*I to give a characteristic restriction profile for *Calanus helgolandicus*, *Calanus finmarchicus*, *Calanus glacilais* or *Calanus hyperboreous*. There were no aberrant restriction profiles that were not characteristic of a particular species.

7.3.3 Comparison of morphological and molecular identification of *Calanus* species from Lurefjorden

Calanus collected from Lurefjorden, following morphological identification in Norway, were characterised to species level using the developed molecular technique. Out of 154 animals, 153 were identified successfully and 1 failed to produce an amplification product. The molecular results showed that 67% of the *Calanus* collected were *C. glacialis*, 32% were *C. helgolandicus*, and 1% were *C. finmarchicus*. No *C. hyperboreous* were found (Figure 7.1). In comparison with determination of species by molecular techniques, 98% of *Calanus* were attributed to *C. helgolandicus* and/or *C. glacialis* species and 2% to *C. finmarchicus* species, based on morphological identification (Figure 7.1). Morphological identification was restricted to stage V copepodites and adult females. The criteria used for the identification did not allow the separation of *C. helgolandicus* and *C. glacialis* species, and no convincing morphological trait was found to classify males to species level. The morphological methods of identification proved ambiguous and even after detailed examination of the morphological characteristics it was noted that many individuals were identified 'with doubt'.

A comparison between the molecular and morphological identification of *Calanus* highlights the limitations of morphological identification methods, and demonstrates

clearly the potential of the molecular technique. The morphological identification was so limited that it precluded a direct comparison of species identification of individuals by morphological and molecular means. The results of the molecular technique showed that most *Calanus* were identified as either *C. helgolandicus* or *C. glacialis*, whereas by morphological means no such differentiation was possible between these two species. Molecular analysis also determined that 1% of the animals were *C. finmarchicus*, compared with 2% based on morphological identification. Only one *Calanus* individual was identified as *C. finmarchicus* by both morphological and molecular means.

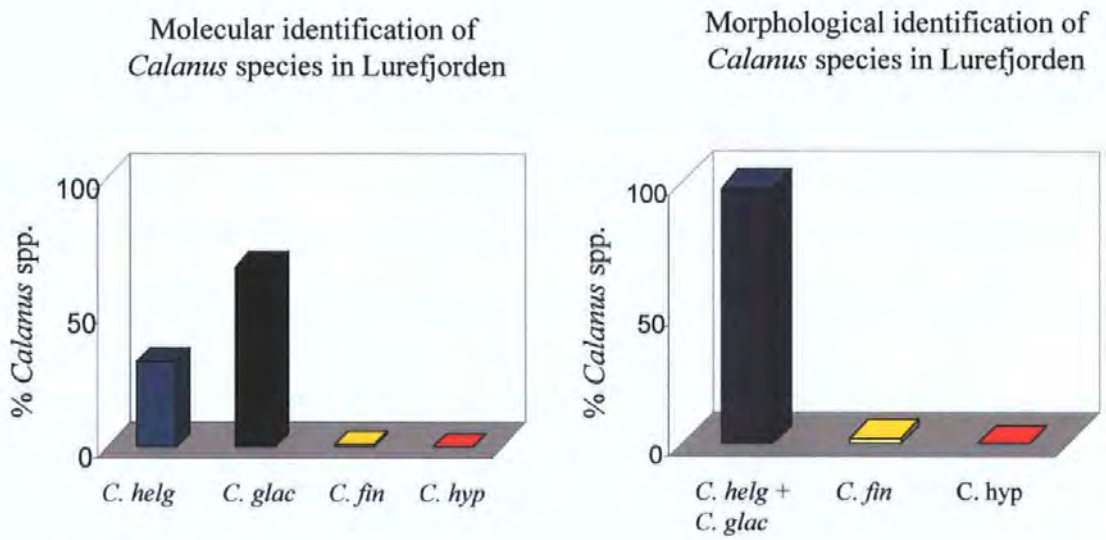


Figure 7.1 Species identification of *Calanus* from Lurefjorden. Comparison of morphological and molecular identification techniques. *C. helg* = *Calanus helgolandicus*, *C. glac* = *Calanus glacialis*, *C. fin* = *Calanus finmarchicus*, *C. hyp* = *Calanus hyperboreous* (n=153).

7.3.4 Molecular identification of *Calanus* species in a mesocosm study

For the mesocosm study in Tromsø, a representative aliquot of 30 females from each of the two cohorts was identified using molecular techniques at the beginning and end of the study. At the beginning of the study, molecular analysis showed the cohort collected from Raunefjorden (Bergen) contained 61% *C. helgolandicus* and 39% *C. finmarchicus*. The cohort collected from Grøtsund (Tromsø) contained 90% *C. finmarchicus* and 10% *C. glacialis* (Figure 7.2). A further 30 individuals, identified from each cohort at the end of the mesocosm study, showed that the Bergen and Tromsø cohort contained 95% *C. finmarchicus* and 5% *C. glacialis* (Figure 7.2). The molecular technique demonstrated clearly the high level of misidentification of females collected from Bergen and consequently, the high level of *C. helgolandicus* in the Bergen cohort during the initial stages of the mesocosm experiment. The molecular identification demonstrated a change in species composition towards a dominance of *C. finmarchicus* in both cohorts during the experiment. Further molecular identification was performed on 20 representative nauplii, collected on day 7 from both cohorts. The species composition of the day 7 nauplii was shown to be 100% *C. finmarchicus* in the Bergen cohort, and 95% *C. finmarchicus* and 5% *C. glacialis* in the Tromsø cohort. This additional analysis demonstrated that the change in species composition, within both cohorts, occurred during the early stages of the mesocosm experiment (Figure 7.2).

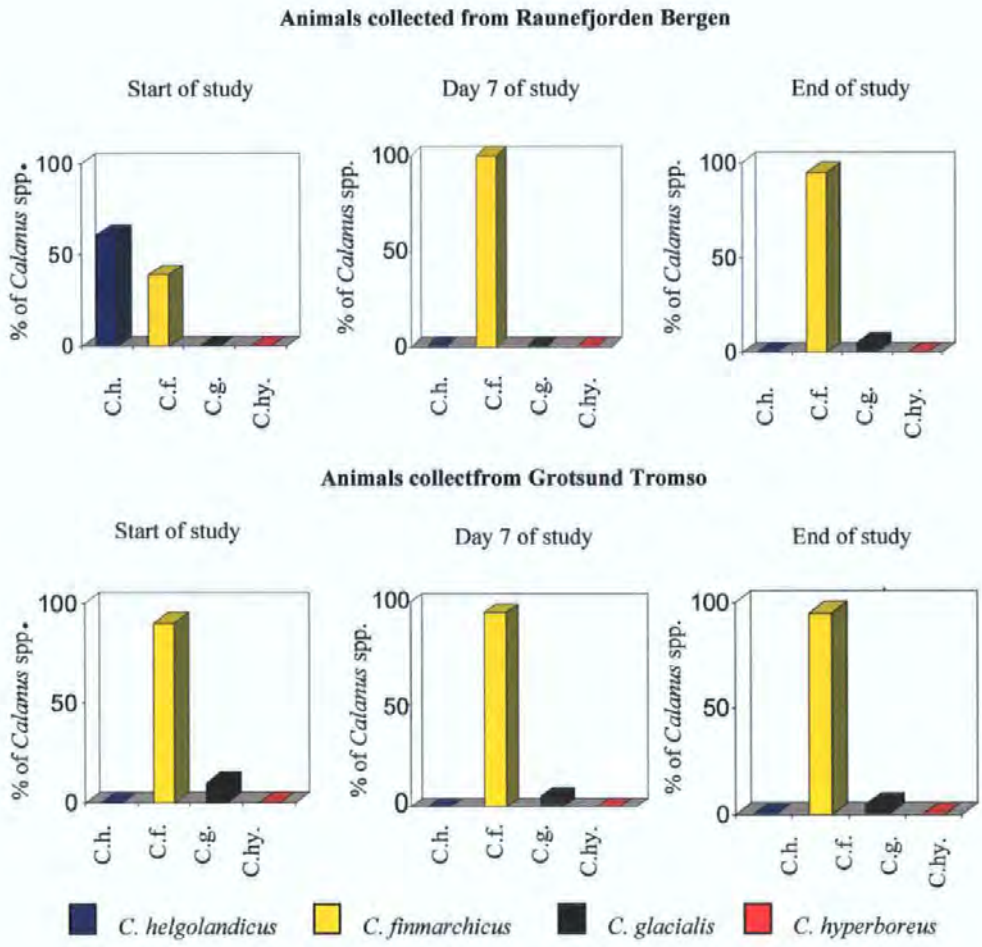


Figure 7.2 The species composition of two *Calanus* cohorts originating from 60°N (Bergen) and 69°N (Tromsø). Representative individuals (20 –30) at the beginning, day 7, and end of the study were identified to species level using molecular techniques.

7.3.5 Distribution of *Calanus* species in the North Atlantic determined by the molecular technique

Between 20 and 30 *Calanus* individuals were characterised to species level, using 16S rDNA amplification and RFLP analysis, from each of 18 different geographical locations within the North Atlantic. (Figure 7.3).

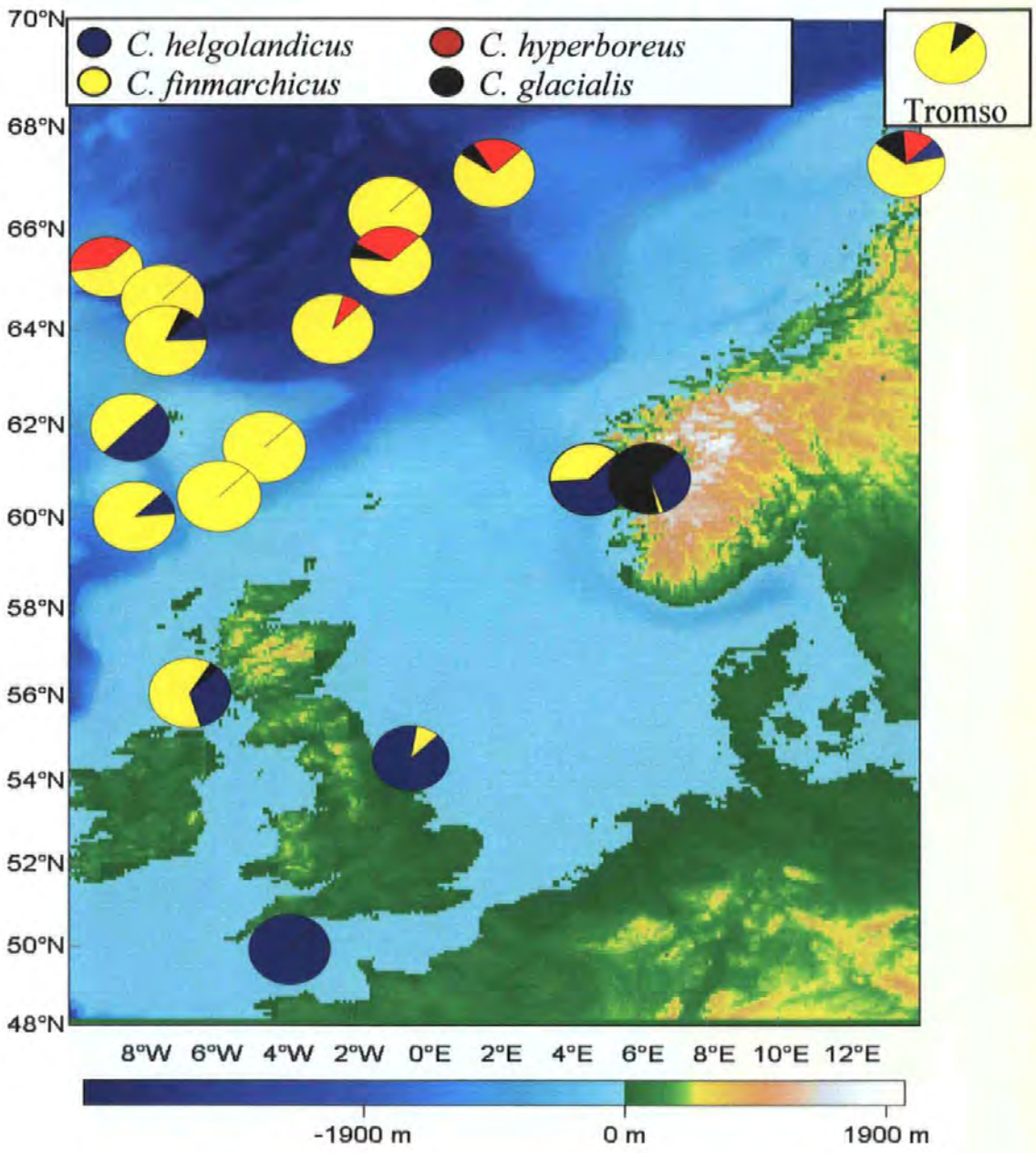


Figure 7.3 Distribution of *Calanus helgolandicus*, *Calanus finmarchicus*, *Calanus glacialis* and *Calanus hyperboreous*, identified using 16S rDNA amplification and RFLP analysis.

Molecular identification has shown that *C. helgolandicus* is distributed predominantly in the south of the area studied, with the species solely representing the composition of *Calanus* in the sample from the English Channel. At 8 of the locations sampled *C. helgolandicus* is found to co-occur with *C. finmarchicus*, and at 4 of these locations, *C. helgolandicus* and *C. finmarchicus* co-occur with *C. glacialis*. *Calanus finmarchicus* has the greatest area of distribution, ranging from the North Sea (55°N), to Tromsø (69°N) and dominating the species composition in the Faroe Shetland channel and the North Atlantic basin. At the most southern range, *C. finmarchicus* co-occurs extensively with *C. helgolandicus*, and in the northern range with *C. glacialis* and/or *C. hyperboreous*. *Calanus glacialis*, although having a wide area of distribution (Oban, 56°N to Tromsø, 69°N), contributes only a small percentage of the *Calanus* species composition. The only location where *Calanus glacialis* dominates is Lurefjorden (Norway). *Calanus hyperboreous* is found in the most northern of the locations sampled, and co-occurs with *C. helgolandicus, C. finmarchicus and C. glacialis.* At one location only are all four *Calanus* species found together (Saltenfjord, 67°N).

7.4 Discussion

The simple molecular identification technique, involving amplification of 16S rDNA and subsequent RFLP analysis, has been applied successfully to a number of samples collected from a wide geographical area. The method was able to distinguish unambiguously between *Calanus helgolandicus*, *Calanus finmarchicus*, *Calanus glacialis* and *Calanus hyperboreous*.

Modifications to the original technique (Chapter 6; Lindeque *et al.*, 1999), included the use of *Taq* DNA polymerase (Promega UK Ltd.) in preference to Dynazyme (Flowgen Instruments Ltd.). *Taq* DNA polymerase was found to give the same amplification success rate as Dynazyme but significantly reduced the cost of analysis. Consequently, with the

use of *Taq* DNA polymerase, which is less thermally stable than Dynazyme, the temperature of the denaturation step of the PCR reaction was reduced from 96°C to 94°C. To reduce the cost of analysis further it was possible to halve the amount of restriction enzyme used (from 5U to 2.5U) for each reaction, with no detrimental effect to the analysis.

Animals stored in absolute ethanol, even at a high concentration of approximately 100 individuals per 1 ml of ethanol, were suitable for DNA amplification. This system of preservation is simple to use at sea, and appears suitable for transportation and long term storage prior to molecular analysis.

The success rate for amplification of the 16S rDNA fragment from preserved animals was over 99%. The small proportion of individuals, less than 1%, that failed to produce an amplification product, can probably be attributed to either insufficient homogenisation of the animal or to insufficient mixing of the PCR reactants prior to allocation to the individual samples. In all cases of successful amplification, the DNA product was digested successfully with *Vsp*I and *Dde*I to give a restriction profile characteristic of one of the four *Calanus* species. There were no aberrant restriction profiles from any geographical location. The application of the developed molecular technique has confirmed that the method can discriminate unambiguously between the four *Calanus* species common to the North Atlantic. Over the geographical region studied the nucleotide sequences at the *Dde*I and *Vsp*I restriction sites are suitably conserved, proving that geographically distant conspecific populations do not affect the method.

An interesting result, revealed by the application of this molecular diagnostic technique, is the species composition of *Calanus* identified in Lurefjorden. Lurefjorden is virtually landlocked and has an invertebrate-dominated predator system. Compared to other fjords, Lurefjorden has a higher standing stock of *Calanus* that exhibit a high mortality rate early in life but a lower mortality rate in the early copepodite stages (Eiane *et*

al., 1999). As with neighbouring fjords, it was thought that the Lurefjorden *Calanus* species composition predominantly included *C. finmarchicus* and *C. helgolandicus.* In contrast, the Lurefjorden species exhibited different egg production rates and predator/prey relationships (Eiane *et al.*, 1999). The application of a molecular technique has revealed the presence of high numbers of *C. glacialis* in the fjord. The molecular-based systematic approach to *Calanus* has proved invaluable, exceeding the capabilities of traditional morphological taxonomy for the *Calanus* species in Lurefjorden. This molecular technique has alerted ecologists to the presence of three *Calanus* species in the fjord and has provided a tool for further research of this unusual ecosystem.

Comparison of the morphological and molecular identification of *Calanus* species in Lurefjorden highlighted the weaknesses of traditional morphological methods. The morphological criteria used for identification were unable to discriminate between male *Calanus*, and did not allow separation of *C. helgolandicus* and *C. glacialis.* Although comparison between morphological and molecular characterisation techniques has obviously been limited, it is clear that a simple method of molecular identification is effectual, being more extensive and less ambiguous than morphological methods.

The molecular method of *Calanus* species discrimination has also proved invaluable for the identification of species in the Tromsø mesocosm experiment. The aim of the mesocosm experiment was to compare growth and mortality rates of two parallel cohorts of *Calanus finmarchicus* originating from 60°N (Bergen) and 69°N (Tromsø). The results of the mesocosm showed that the Bergen cohort had the same overall life cycle as the Tromsø cohort, suggesting that the generation shifts of *Calanus finmarchicus* are environmentally governed. However, a distinct difference between the two cohorts was the high instantaneous mortality exhibited in the first nauplii stages of the Bergen cohort (Andreassen *et al.*, 1999). The mortality was not due to food availability, since both the mesocosms had similar food compositions and concentrations. Not until the molecular

identification of animals at the beginning of the study demonstrated that 61% of the Bergen cohort was in fact *C. helgolandicus*, was it presumed that the mortality rate could be ascribable to species specific differential mortality. Further analysis of nauplii at day 7 in both cohorts, only identifiable to species level by molecular techniques, showed a shift in species composition, confirming that the high mortality in the Bergen sample was indeed attributed to *C. helgolandicus*. Only with the molecular identification technique was it possible to conclude that the high instantaneous mortality rate was species specific, probably ascribable to the temperature shock of NI *C. helgolandicus*.

Molecular identification of *Calanus* species from a wide geographical area within the North Atlantic, has not only proved the accuracy and reproducibility of the technique, but has also stimulated some new interpretations of dispersal. The distribution of *Calanus* species determined by molecular analysis is mostly reflective of traditional results. However, the molecular analysis has clearly shown extended areas of distribution and co-occurrence. These results demonstrate the limitations of identification of *Calanus* species based solely on morphology and have highlighted the potential errors associated with relying on geographical location for species determination. *Calanus helgolandicus* is a warm-temperate water species found in southern areas of the North Atlantic, the Celtic Sea, North Sea and in coastal waters south of England (Fleminger and Hulsemann, 1977; Planque and Fromentin, 1996). The occurrence of *C. helgolandicus* determined by this molecular study is generally supportive of the traditional distributions described above. It is however, surprising to find *C. helgolandicus* north of the Faroes, especially considering that the sample was collected in April when the water temperature was still low. Possibly, *C. helgolandicus* were found in this area due to advective transport in the Atlantic stream.

Calanus finmarchicus is more a cold-temperate water species than *C. helgolandicus*. The species is abundant in coastal and Atlantic water around Iceland (Gislasson and Astthorsson, 1995), and is found in the East Icelandic current water and the

Arctic intermediate water around the Faroes. In this study, *C. finmarchicus* was observed as far south as 54°N in the North Sea and as far north as Grotsund, Tromsø. Traditionally *C. finmarchicus* is thought to overlap extensively with *C. helgolandicus* in many areas of the North Atlantic and in the shelf seas around the United Kingdom (Williams and Conway, 1980). The two *Calanus* species co-occur extensively in surface waters of the Northeast Atlantic during spring, and in the surface waters of the North Sea during summer and autumn (Planque and Fromentin, 1996). The molecular analysis of samples clearly reflects this trend, with a co-occurrence of *C. finmarchicus* and *C. helgolandicus* seen in the Faroe and west of Scotland samples collected in April and May, and in the North sea sample obtained in July. Planque and Fromentin (1996), from an analysis of CV and CVI animals collected in the subsurface layer, estimated that the region of co-occurrence of *C. finmarchicus* and *C. helgolandicus* extended from 52°N to 57°N in the North Sea and Northeast Atlantic. It is obvious from this molecular study that the sympatric area of these two species extends well above 57°N. Compared with traditional results, it is certainly unusual to see *C. helgolandicus* co-occurring at latitudes of 64°N.

Calanus glacialis, predominantly an Arctic water species (Frost, 1974), has a reproductive range overlapping with both *C. helgolandicus* and *C. finmarchicus* (Fleminger and Hulsemann, 1977). From the molecular analysis of samples there are four regions where these three species of *Calanus* co-occur: north of the Faroe isles; Saltenfjord and Lurefjorden, western Norway; and Oban, western Scotland. It was perhaps, unanticipated to see *C. glacialis* as far south as Oban compared to traditional, morphologically determined distributions.

Calanus hyperboreous is a northern Arctic cold water species that has a population centred in the Greenland Sea (Conover, 1988; Hirche, 1991). *Calanus hyperboreous* shows overlapping distributions with *C. finmarchicus* and *C. glacialis* at the front between the Atlantic and Arctic water masses in the Northeast Atlantic (Hirche *et al.*, 1994).

Molecular identification of *Calanus* has shown three sympatric areas of *C. hyperboreous*, *C. glacialis* and *C. finmarchicus*.

This molecular study has demonstrated a greater co-occurrence of *Calanus* species than traditionally thought. The extent of the sympatric areas may not have been previously realised, because geographical location has played a major role in species determination, with obvious consequences. Taking into account the difficulties associated with morphologically distinguishing between *Calanus* species, especially considering the extent of overlap in the length distributions of *C. finmarchicus* and *C. glacialis* (Jaschnov, 1972; Frost, 1974), it is not unfair to assume that *Calanus* species have been misidentified previously, prejudicing prescribed geographical distributions. Traditionally, descriptions of *Calanus* distributions have been based on CV and CVI animals, because of the difficulties in identifying earlier stages. Williams and Conway (1980) gave evidence that the vertical temperature gradient was responsible for differences in the vertical distributions of co-occurring *C. finmarchicus* and *C. helgolandicus*, and between different stages within a species. However, because of the difficulty in identifying early developmental stages, assumptions on the vertical distribution of different stages from a site of sympatric distribution had to be drawn from analysis of isolated species. The developed molecular tool, allowing the unambiguous identification of *Calanus* species at any developmental stage, has obvious implications in describing the distributions of *Calanus* juveniles. In areas of extreme distribution, or unexpected co-occurrence, it is not known whether adult populations are coincident with younger populations. Such a question is of great ecological importance because the presence of young stages would be indicative of a growing population. If no eggs or nauplii were present, it may be assumed that the older stages have been transported to their site of collection. Answering such questions is only possible with the developed molecular technique.

The four *Calanus* species, for which this molecular identification technique has been designed, are the predominant species in the North Atlantic. However, *Calanus marshallae* has been identified recently in Arctic waters in Isfjorden, Spitsbergen (Sundt and Melle, 1998) by sequence data analysis of an amplified region of 16S rDNA. This appears to be an isolated case and therefore, does not warrant extension of the developed technique to include such species.

The relatively simple and inexpensive molecular method, developed for the identification of the North Atlantic *Calanus* species, is reliable and reproducible. The technique has proved invaluable in helping to interpret unexplained biological differences between *Calanus* populations and has shown a huge potential for better describing and understanding the distribution of this important copepod genus. Correct species identification is the keystone of ecological studies, and accurate descriptions of the spatial and temporal patterns of *Calanus* species are invaluable for understanding and predicting the response of the different species to environmental conditions.

CHAPTER 8

Final Discussion and Future Work

CHAPTER 8

Final Discussion and Future Work

In comparison with other areas of biological research, the application of molecular techniques to biological oceanography is somewhat in its infancy (Section 1.2.2), despite the potential of such methods to increase our understanding of many fundamental marine biological processes. In this study, molecular techniques such as bacteriophage cloning, genomic library construction, IVPCR, mRNA analysis (using reverse transcriptase PCR), and DNA sequence analysis, have all been successfully applied to the study of two aspects of *Calanus* ecology. The processes of *C. helgolandicus* production and recruitment, and the closely related problem involving species identification, have thus begun to be addressed.

The vital role played by developmental genes in the pathway whereby a fertilised egg develops into an adult make them not only of considerable scientific interest, but also provides a potential tool with which to address measurements of production and recruitment. For the first time, sequence data has been obtained from genes implicated in playing critical roles in *C. helgolandicus* development. This sequence data from *C. helgolandicus* developmental genes provides a means by which probes can be produced to investigate their temporal expression. In particular, appropriate probes could provide excellent tools with which to estimate key production processes, such as copepod egg viability. To gain a full understanding of *Calanus* population dynamics it is imperative that egg production measurements are accompanied by a measurement of egg hatching success. Such a study is essential, since variability of copepod recruitment may depend largely on the rate of production of viable eggs rather than on fecundity *per se* (Miralto *et al.*, 1999). *C. helgolandicus* developmental genes have also provided entirely novel tools

with which to investigate *Calanus* developmental mechanisms. Analysis of sequence data from such genes provides confirmation that the genetics of *C. helgolandicus* development is consistent with the conserved nature of developmental pathways throughout the animal kingdom.

The first specific *C. helgolandicus* developmental gene sequence was acquired by probing the constructed *Calanus helgolandicus* genomic library with a previously cloned homeobox from *C. helgolandicus*. The gene fragment has been named *Cal-Antp* based on its homology with the *Drosophila* homeotic gene *Antennapedia*. The *Cal-Antp* gene appears to be a direct homologue, not only of the *Drosophila Antennapedia* gene, but also of the *Artemia* (Brine shrimp) and *Ethmostigmus* (Centipede) *Antennapedia* genes. This homology supports previous comparative studies, which have indicated that crustaceans and insects share almost identical complements of *Hox* genes. In particular, the homeotic genes that specify middle body regions, such as *Antennapedia*, are known to have originated before the divergence of the insect and crustacean lineages (Averof and Akam, 1993), and have consequently maintained a high degree of homology.

The RT-PCR system developed has shown the *C. helgolandicus Antennapedia* homologue to be expressed in a temporally specific manner, from eggs greater than 18 hours old through to copepodite stages. The *Cal-Antp* gene was not detected in eggs less than 18 hours old or in adult animals. The *Antennapedia* gene is a homeotic selector gene, these collectively being responsible for the antero-posterior axis of the body plan through their roles as transcription factors. Considering the remarkable conservation of function demonstrated by homeotic genes, it is possible to surmise that *Cal-Antp* plays an important role in the definition of the *Calanus* body plan. *Cal-Antp* is the first developmental gene to be analysed in *Calanus*, and has opened new avenues through which to explore the reproductive and developmental biology of this important copepod. The duration of expression of *Cal-Antp* does not make it ideal for use as an egg viability probe for use on

eggs taken directly from the field, since expression in the egg is detected only after 18 hours. However, the gene has exciting potential for use as an egg viability probe in laboratory studies.

Three further homeobox-containing genes have successfully been identified and analysed through their expression in *C. helgolandicus* eggs. Two of these genes are *caudal* homologues, and the third is homologous to the *Antennapedia* class of genes. The role of *caudal* genes in other organisms has been shown to be involved in the initial establishment of morphogenic gradients. In contrast to many cases including *Drosophila,* where only one *caudal* gene is present, it is interesting to note the presence of two *caudal* gene homologues in *C. helgolandicus.* Such duplication has only previously been reported in vertebrate systems. Further analysis and mapping of these *caudal* genes on the *Calanus* genome, although considerably outside the realms of this study, may provide some interesting insights of evolutionary significance. The *C. helgolandicus Antennapedia* class homologue, termed clone12-RTPCR in this study, is a particularly exciting candidate for future use as an egg viability probe. This gene shows a high degree of homology with the sex combs reduced (*scr*) gene of *Drosophila* (an A*ntennapedia* class gene) isolated from 0-22 hour embryos (LeMotte *et al.*, 1989), and therefore warrants further investigation to determine the duration of expression in *C. helgolandicus* embryos.

A detection system for one or more such genes involved in the early stages of embryogenesis would enable large numbers of eggs to be simultaneously analysed via their mRNA content to determine whether transcription of the gene(s) in question has been initiated. Detection of this mRNA would be indicative of a normal developmental pathway being followed, presumably leading to hatching and naupliar development. If the transcription initiation of such developmental genes is not detected it would be assumed that the eggs were not undergoing normal development and were therefore non-viable. Whilst the development of a probe to such a gene would provide a remarkable tool for such

analysis, the entire technique will further require a system for fixation and permeabilisation of the eggs such that probes can be hybridised to mRNA within the cells. Background hybridisation to genomic DNA will be negligible, as such genes are normally present at only one copy per cell, whereas the developmental gene transcript should be expressed at a relatively high level in developing embryos.

The development of a molecular system with which to ascertain the viability of *Calanus* eggs could be applied to greatly increase our understanding of the factors affecting *Calanus* recruitment. A reliable and accurate measurement of egg viability would be invaluable to quantify the input of young nauplii into the population, in turn allowing a more accurate assessment of mortality rates than those predicted purely from female fecundity. Another particularly important application of such a technique would be its use in the current 'diatom debate'. It has been observed in many copepod species that hatching success of newly spawned eggs varies between 0-100% (Poulet *et al.*, 1994; 1995; Ianora *et al.*, 1995). Recent laboratory studies have indicated that the production of inhibitory compounds by diatoms has a negative effect on egg hatching success (Poulet *et al.*, 1994; Miralto *et al.*, 1995; Ianora *et al.*, 1995; Chaudron *et al.*, 1996; Ianora *et al.*, 1996; Uye, 1996: Ban *et al.*, 1997; Miralto *et al.*, 1999). However, the importance of these studies is still controversial, as it is unclear whether the diatom effect is due to toxicity or low nutritional value (Jónasdóttir *et al.*, 1998; Ianora *et al.*, 1999). The debate is of particular interest since if the diatoms do possess an inhibitory compound, suggested by Ianora *et al.* (1999) to be specific aldehydes, then one would expect to find a decrease in egg viability when diatoms are abundant in the field. On the other hand, if it is a case of low nutritional value, copepods in the field may compliment their diet with other components of microplankton. At present there is no method of measuring the viability of *in situ* eggs without a period of laboratory incubation and thus, the consequence of artefacts. The potential to ascertain the viability of eggs taken directly from the field

utilising molecular probes, would provide a powerful tool with which to increase understanding of the key production process of copepod egg viability.

Perhaps the greatest potential of an egg viability probe is apparent when combined with techniques such as flow cytometry. Single or multiple probes could be suitably labelled such that detection of expression of the specific gene(s) indicative of a normal developmental pathway being followed could be coupled with species specific probes targeted towards cytoplasmic ribosomal RNA. Such combined species-specific and egg viability measurements of *Calanus* could thus become a routine, even ship based, technique.

The 'diatom debate' has also highlighted a second important potential use for *C. helgolandicus* developmental gene sequences. The development of RT-PCR expression assays for a number of *C. helgolandicus* developmental genes would permit the effect of environmental factors, such as nutritional status, on the molecular aspects of development to be addressed. Laboratory studies have shown that the inhibitory compounds within diatoms are accumulated in *Calanus* oocytes during oogenesis. The inhibitory compounds arrest embryonic development and are thought to cause two major abnormalities during mitosis; mismatch between nuclear division and cytokinesis; and an absence of cell membranes in daughter cells (Poulet *et al.*, 1994, 1995; Miralto *et al.*, 1999). The mechanisms by which toxins, such as the suggested aldehydes, inhibit embryogenesis are as yet unknown. *C. helgolandicus* developmental genes acquired would be of considerable interest for the development of RT-PCR expression assays. Determining the expression of these key developmental genes would provide the first understanding at the fundamental biological level of the effects of inhibitory compounds on *Calanus* embryogenesis and developmental mechanisms.

A closely related problem in analysing copepod production and recruitment involves the positive identification of *Calanus* to species level. Diagnostic morphological

characteristics of *Calanus* species are restricted essentially to minor variations in their secondary sex characteristics, presenting a persistent problem in the identification of individuals, especially at immature stages. In view of these difficulties experienced by *Calanus* ecologists, part of this study has involved the development of a relatively simple and cost effective method, using molecular techniques, with which to unambiguously identify the four *Calanus* species common to the North Atlantic. The developed system involves the PCR amplification of a region of the mitochondrial 16S rRNA gene without prior purification of the DNA, followed by Restriction Fragment Length Polymorphism (RFLP) analysis of the amplified product. Using this system it has been possible to identify not only whole animals at any developmental stage from egg to adult, but also to identify individual body parts to species level. Such identification capabilities for body parts have significant implications for predator gut content analyses in feeding behaviour studies.

The molecular identification technique has been applied to field samples and has shown the unexpected presence of three different *Calanus* species in Lurefjorden, Norway. Morphological indicators for the identification of individual *Calanus* species in Lurefjorden proved insufficient to distinguish accurately between any of the species, especially for distinction between *Calanus helgolandicus* and *Calanus glacialis* individuals. The presence of three *Calanus* species in the fjord, as indicated using molecular identification techniques, in contrast to the expected one or two, has alerted the *Calanus* ecologists to the diversity of species' in the fjord ecosystem, and has helped to explain the unusual population dynamics. Unexplained biological events, such as the high initial mortality rate in one cohort of the Tromsø mesocosm, have also been clarified by the use of molecular methods, proving the application and versatility of such techniques.

The technique developed has also been successfully applied to a number of samples collected from a wide geographical area in the North Atlantic. There were no aberrant

restriction profiles from any of the geographical locations studied, confirming that the method was consistently accurate and was not affected by geographically distant conspecific populations. The study additionally demonstrated an extended area of distribution and a greater level of co-occurrence of some species than traditionally thought. The distribution of *Calanus* species, determined by molecular analysis, has highlighted the limitations of morphological identification and the potential errors associated with relying on geographical location as a factor in species determination. In ecological terms it would be of considerable interest to determine whether unexpected cases of occurrence, such as *Calanus glacialis* in coastal waters off Oban (Scotland), and the presence of *Calanus helgolandicus* at 64°N, are coincident with a young population. That is, whether these *Calanus* species at the limit of their distribution are solely found as a consequence of advection or whether the species are capable of reproducing at such locations. An accurate distribution of young *Calanus* developmental stages would also allow us to ascertain whether sympatric species result from an interleaving of water masses or whether the species are reproductively active in both water masses. This important ecological question can only be answered utilising the powerful molecular method capable of determining species identity at any developmental stage.

The molecular-based approach to *Calanus* species determination is the first technique to allow identification of all developmental stages, and thus has huge implications for the understanding of population dynamics. Such knowledge is needed for predictive capabilities of the abundance and dynamics of populations and their response to environmental and anthropogenic factors. The developed method is likely to find an important niche in the research of this genus. Accurate species identification is the keystone to any ecological study and there are many opportunities for increasing understanding of *Calanus* life histories, ecological roles and population dynamics, including reproductive processes, using such a technique.

There is considerable potential for the use of molecular techniques in copepod systematics. Expansion of the species identification system to other copepods found in British waters, for which identification of either adults or earlier developmental stages is problematic, is of worth. Another application of such molecular techniques within the realm of marine copepods is the categorizing of undescribed juveniles. In obtaining 16S rDNA sequence data from juvenile animals (after intricate recording of their morphology), it should be possible in some cases to link hitherto undescribed juvenile forms to adults of a known species. This would increase knowledge of the early life stages of copepods for which morphology alone cannot be used to link the juvenile and adult life stages.

Calanus are important members of the zooplankton community and only by understanding the fundamental biological processes of development, the factors affecting them, and by accurately describing the temporal and spatial distribution of all *Calanus* stages, can we hope to increase our predictive capabilities. This study has bought powerful molecular techniques into *Calanus* ecology, and has clearly shown the potential of such methods to increase understanding of *Calanus* development and systematics.

REFERENCES

Abouheif, E., 1997. Developmental genetics and homology: a hierarchial approach. TREE, **12**, 405-408.

Akam, M., 1989. Hox and HOM: homologous gene clusters in insects and vertebrates. Cell, **57**, 347-349.

Ambler, J. W., 1985. Seasonal factors affecting egg production and viability of eggs of *Acartia tonsa Dana* from East Lagoon, Galveston, Texas. Est. Coast. Shelf. Sci., **20**, 743-760.

Andreassen, P., Marker, T., Hansen, B. W., and Tande, K. S., 1998. Growth and mortality rates in two populations of *Calanus finmarchicus* originating from 60°N and 69°N: Mesocosm study in Tromsø, Norway, during spring 1998. Poster presentation, third annual meeting of TASC, Villefrance, France, January, 1999.

Averof, M., and Akam, M., 1993. HOM/Hox genes of *Artemia*: implications for the origin of insect and crustacean body plans. Curr. Biol., **3**, 73-78.

Averof, A., and Patel, N. H., 1997. Crustacean appendage evolution associated with changes in *Hox* gene expression. Nature, **388**, 682-686.

Ban, S., Burns, C., Castel, J., Chaudron, Y., Christou, E., Escribano, R., Umani, S. F., Gasparini, S., Ruiz, F. G., Hoffmeyer, M., Ianora, I., Kang, H.-K., Laabir, M., Lacoste, A., Miralto, A., Ning, X., Poulet, S., Rodriguez, V., Runge, J., Shi, J., Starr, M., Uye S.-I., and Wang, Y., 1997. The paradox of diatom-copepod interactions. Mar. Ecol. Prog. Ser., **157**, 287-293.

Beckman, B. R., and Peterson, W. T., 1986. Egg production by *Acartia tonsa* in Long Island Sound. J. Plankton Res., **8**, 917-925.

Bergeron, J. P., and Alayse-Danet, A., M., 1981. Aspartate transcarbamylase de la coquille sait Jacques *Pecten maximus* L. (Mollusque Lamellibranche): méthode de dosage et variations de l'activité dans le manteau et la gonade. J. Exp. Mar. Biol. Ecol., **50**, 99-117.

Bergeron, J. P., 1982. L'Aspartate transcarbamylase, indice de croissance des organismes marins: perspectives et limities. Publ. CNEXO (Actes Colloq.), **14**, 177-192.

Bergeron, J. P., 1995. Aspartate transcarbamylase activity in the assessment of mesozooplankton production: new aspects from oceanic areas. ICES J. Mar. Sci., **52**, 305-313.

Berggreen, U., Hansen, B., and Kiørboe, T., 1988. Food size spectra, ingestion and growth of the copepod *Acartia tonsa* during development; implications for determination of copepod production. Mar. Biol., **99**, 341-352.

Biegala, I. C., Harris, R. P., and Bergeron, J. P., 1999. ATCase activity, RNA:DNA ratio, gonad development stage, and egg production in the female copepod *Calanus helgolandicus*. Mar. Biol., **135**, 1-10.

Britschgi, T. B., and Giovannoni, S. J., 1991. Phylogenetic analysis of a natural marine bacterioplankton population by rRNA gene cloning and sequencing. App. Envir. Micro., **57**, 1707-1713.

Brooke, N. M., Garcia-Fernàndez, J., and Holland, P. W. H., 1998. The Parahox gene cluster is an evolutionary sister of the *Hox* gene cluster. Nature, **392**, 920-922.

Brown, T. A., 1995. How to obtain a clone of a specific gene, in *Gene Cloning, an introduction, 3rd edition* (ed. T. A. Brown). Chapman and Hall, London. 159-183.

Brown, T. A., 1995. Strategies for gene isolation and analysis, in *Essential molecular biology, volume* 2 (ed. T. A. Brown), IRL press at Oxford, University press, Oxford. 1-13.

Bucklin, A., and Marcus, N. H., 1985. Genetic differentiation of populations of the planktonic copepod *Labidocera aestiva*. Mar. Biol., **84**, 219-224.

Bucklin, A., and Wiebe, P., 1986. Genetic heterogeneity in euphasiid populations: *Euphausia krohnii* and *Nematoscelis megalops* in North Slope water. Limnol. Oceanogr., **31**, 1346-1352.

Bucklin, A., Rienecker, M. M., and Mooers, C. N. K., 1989. Genetic tracers of zooplankton transport in coastal filaments off Northern California. J. Geophys. Res., **94**, 8277-8288.

Bucklin, A., 1991. Population genetic responses of the planktonic *copepod Metridia pacifica* to a coastal eddy in the California current. J. Geophys. Res., **96**, 14,799-14,808.

Bucklin, A, Frost, B. W., and Kocher, T. D., 1992. DNA sequence variation of the mitochondrial 16S rRNA in *Calanus* (Copepoda; Calanoida): intraspecific and interspecific patterns. Mol. Mar. Biol. Biotech., **1**, 397- 407.

Bucklin, A., and LaJeunese, T. C. 1994. Molecular genetic variation of *Calanus pacificus* (Copepoda; Calanoida): preliminary evaluation of genetic structure and sub-specific differentiation based on mtDNA sequences. Calif. Coop. Ocean Fish Invest. Rep., **35**, 45-51.

Bucklin, A., Frost, B. W., and Kocher, T. D. 1995. Molecular systematics of six *Calanus* and three *Metridia* species (Calanoida; Copepod). Mar. Biol., **121**, 655-664.

Bucklin, A., Sundt, R. C., and Dahle, G. 1996. The population genetics of *Calanus finmarchicus* in the North Atlantic. Ophelia, **44**, 29-45.

Bucklin, A., and Kocher, T. D., 1996. Source regions for recruitment of *Calanus finmarchicus* to Georges Bank; evidence from molecular population genetic analysis of mtDNA. Deep-Sea Res., **43**, 1665-1681.

Bucklin, A., Bentley, A. M., and Franzen, S. P., 1998. Distribution and relative abundance of *Pseudocalanus moultoni* and *P. newmani* (Copepoda: Calanoida) on Georges Bank using molecular identification of sibling species. Mar. Biol., **132**, 97-106.

Bucklin, A., 1998. In: Proceedings of the Second International Conference on pelagic Biogeography, 9-14 July. The Netherlands. (eds. Pierrot-Bults, A. C., and van der Spoel, S.) UNESCO Technical Papers in Marine Science.

Bucklin, A., and Wiebe, P. H., 1998. Low mitochondrial diversity and small effective population sizes of the copepods *Calanus finmarchicus* and *Nannocalanus minor*: possible impact of climate variation during recent glaciation. The Amer. Genetic Assoc., **89**, 383-392.

Bürglin, T. R., Ruvkun, G., Coulson, A., Hawkins, N. C., McGhee, J. D., Schaller, D., Wittman, D., Müller, F., and Waterston, R. H., 1991. Nematode homeobox cluster. Nature, **351**, 703.

Bürglin, T. R., 1994. A comprehensive classification of homeobox genes, in *Guidebook to the Homeobox genes* (ed. D. Duboule), Oxford University Press, Oxford, 25-73.

Capovilla, M., Eldon, E. D., and Pirrotta, V., 1992. The *giant* gene of *Drosophila* encodes a b-ZIP DNA-binding protein that regulates the expression of other segmentation genes. Dev., **114**, 99-112.

Carroll, S. B., 1995. Homeotic genes and the evolution of arthropods and chordates. Nature, **376**, 479-485.

Casares, F., and Mann, R. S., 1998. Control of antennal versus leg development in *Drosophila*. Nature, **392**, 723-726.

Chaudron, Y., Poulet, S. A., Laabir, M., Ianora, A., and Miralto, A., 1996. Is hatching success of copepod eggs diatom density-dependent? Mar. Ecol. Prog. Ser., **144**, 185-193.

Checkley, D. M., Dagg, M. J., and Uye, S.-I., 1992. Feeding, excretion and egg production by individual and populations of the marine, planktonic copepods, *Acartia spp.*, and *Centropages furcatus*. J. Plankton Res., **14**, 71-96.

Clarke, L., and Carbon, J., 1976. A colony bank containing synthetic Col El hybrid plasmids representative of the entire *E. coli* genome. Cell, **9**, 91-99.

Cohen, S. M., Brönner, G., Kütter, F., Jürgens, G., Jäckie, H., 1989. *Distal-less* encodes a homeodomain protein required for limb development in *Drosophila*. Nature, **338**, 432-434.

Colebrook, J. M., 1982. Continuous plankton records: seasonal variations in the distribution and abundance of plankton in the North Atlantic Ocean and the North Sea. J. Plankton Res., **4**, 435-462.

Colebrook, J. M., 1985. Continuous Plankton Records: over-wintering and annual fluctuation in the abundance of zooplankton. Mar. Biol., **84**, 261-265.

Collins, J., and Hohn, B., 1978. Cosmids: A type of plasmid gene-cloning vector that is packageable *in vitro* in bacteriophage λ heads. Proc. Natl. Acad. Sci., USA, 75, 4242.

Conover, R. J., 1988. Comparative life histories in the genera *Calanus* and *Neocalanus* in high latitudes of the northern hemisphere. Hydrobiologia, **167/168**, 127-142.

Conover, R. J., Wilson, S., Harding, G. C. H., and Vass, W. P., 1995. Climate, copepods and cod: some thoughts on the long-range prospects for a sustainable northern cod fishery. Climate Research, **5**, 69-82.

Conway, D. V. P., McFadzen, I. R. B., and Tranter, P. R. G., 1994. Digestion of copepod eggs by larval turbot *Scophthalmus maximus* and egg viability following gut passage. Mar. Ecol. Prog. Ser., **106**, 303-309.

Corkett, C. J., and McLaren, I. A., 1978. The biology of *Pseudocalanus*. Adv. Mar. Biol., **15**, 1-231.

Curtis, D., Lehmann, R., and Zamore, P. D., 1995. Translational regulation in development. Cell, **81**, 171-178.

Dagg, M. J., and Littlepage, J. L., 1972. Relationships between growth rate and RNA, DNA, protein and dry weight in *Artemia salina* and *Euchaeta elongata*. Mar. Biol., **17**, 162-170.

Damen, W. G. M., Hausdorf, M., Seyfarth, E.-A., and Tautz, D., 1998. A conserved mode of head segmentation in arthropods revealed by the expression pattern of *Hox* genes in a spider. Proc. Natl. Acad. Sci. U. S. A., **95**, 10665-10670.

Degnan, B. M., Degnan, S. M., Giusti, A., and Morse, D. E., 1995. A *hox/hom* homeobox gene in sponges. Gene, **155**, 175-177.

DeLong, E. F., Wickham, G. S., and Pace, N. R., 1989. Phylogenetic stains: ribosomal RNA-based probes for the identification of single cells. Science, **24**, 1360-1363.

DeLorenzi, M., Ali, N., Saari, G., Henry, C., Wilcox, M., and Bienz, M., 1988. Evidence that the *Abdominal-B* r element function is conferred by a trans-regulatory homeoprotein. EMBO J., **10**, 3223-3232.

Devereux J., Haeberli P., and Smithies O., 1984. A comprehensive set of sequence-analysis programs for the VAX. Nucl. Acids Res., **12**, 387-395.

Dickson, R. R., Kelly, P. M., Colebrook, J. M., Wooster, W. S., and Cushing, D. H., 1988. North winds and production in the eastern North Atlantic. J. Plankton Res., **10**, 151-169.

Dickson, R. R., and Brander, K. M., 1993. Effects of a changing windfield on cod stocks of the North Atlantic. Fish. Oceanogr., **2**, 124-153.

Dubnau, J., and Struhl, G., 1996. RNA recognition and translational regulation by a homeodomain protein. Nature, **379**, 694-699.

Duboule, D., and Dollé, P., 1989. The structural and functional organisation of the murine HOX gene family resembles that of *Drosophila* homeotic genes. EMBO J., **8**, 1497-1505.

Duboule, D., 1994. Guidebook to the homeobox genes, (ed. Duboule, D.), Oxford University Press Inc., New York.

Duboule, D., 1998. *Hox* is in the hair: a break in colinearity? Genes and Dev., **12**, 1-4.

Duprey, P., Chowdhury, K., Dressler, G. R., Balling, R., Simon, D., Guenet, J.-L., and Gruss, P., 1988. A mouse gene homologous to the *Drosophila* gene *caudal* is expressed in epithelial cells from the embryonic intestine. Genes and Dev., **2**, 1647-1654.

Eiane, K., Aksnes, D. L., Ohman, M. D., Wood, S., and Martinussen, M., B., 1999. Stage specific mortality of *Calanus spp.* under different predator regimes. Abstract, ICES symposium: Population dynamics of *Calanus* in the North Atlantic – results from the trans-Atlantic study of *Calanus finmarchicus*.

Falkowski, P. G., and Owens, T. C., 1982. A technique for estimating phytoplankton division rates by using a DNA-binding fluorescent dye. Limnol. Oceanogr., **27**, 776-782.

Falkowski, P. G. and LaRoche, J., 1991. Molecular Biology in Studies of Ocean Processes. Int. rev. of Cytology, **128**, 261-303.

Feder, J. N., Lanying, L., and Jan Y. N., 1994. Genomic cloning and chromosomal localisation of *HRY*, the human homolog to the *Drosophila* segmentation gene, *hairy*. Genomics, **20**, 56-61.

Feinberg, A. P., and Vogelstein, B., 1983. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem., **132**, 6-13.

Feinberg, A. P., and Vogelstein, B., 1984. A technique for radiolabelling DNA restriction endonuclease fragments to high specificity. Addendum. Anal. Biochem., **137**, 266-268.

Finkelstein, R., Smouse, D., Capaci, T. M., Spradling, A. C., and Perrimon, N., 1990. The *orthodenticle* gene encodes a novel homeodomain protein involved in the development of the *Drosophila* nervous system and ocellar visual structures. Genes and Dev., **4**, 1516-1527.

Fleminger, A., and Hulsemann, K., 1977. Geographical range and taxonomic divergence in North Atlantic *Calanus* (*C. helgolandicus, C. finmarchicus*, and *C. glacialis*). Mar. Biol., **40**, 233-248.

Fritz, A., and De Robertis, E. M., 1988. *Xenopus* homeobox-containing cDNAs expressed in early development. Nucl. Acids Res., **16**, 1453-1469.

Fromentin, J. –M., and Planque, B., 1996. *Calanus* and environment in the eastern North Atlantic II. Influence of the North Atlantic Oscillation on *C. finmarchicus* and *C. helgolandicus*. Mar. Ecol. Prog. Ser., **134**,111-118.

Frost, B. W., 1974. *Calanus marshallae*, a new species of Calanoid copepod closely related to the sibling species *C. finmarchicus* and *C. glacialis*. Mar. Biol., **26**, 77-99.

Frumkin, A., Rangini, Z., Ben-Yehuda, A., Gruenbaum, Y., and Fainsod, A., 1991. A chicken *caudal* homologue, *CHox-cad*, is expressed in the epiblast with posterior localisation and in early lineage. Dev., **112**, 207-219.

Gannon, F. and Powell, R., 1995. Construction of recombinant DNA molecules, in *Essential molecular biology, volume 1* (ed. Brown, T. A.). IRL press at Oxford, University press, Oxford. 143-178.

Gaunt, S. J., Sharpe, P. T., and Duboule, D., 1988. Spatially restricted domains of homeo-gene transcripts in mouse embryos: Relation to a segmented body plan. Dev. (**suppl. 104**), 169-179.

Gehring, W. J., 1987. Homeo Boxes in the study of development. Science, **236**, 1245-1252.

Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G., and Wüthrich, K., 1994. Homeodomain-DNA recognition. Cell, **78**, 211-223.

Gislasson, A., and Astthorsson, O. S., 1995. Seasonal cycle of zooplankton southwest of Iceland. J. Plankton Res., **17**, 1959-1976.

Govind, S., and Steward, R., 1991. Dorsoventral pattern-formation in *Drosophila*-signal transduction and nuclear targeting. Trends in Gen., **7**, 119-125.

Graham, A., Papalopulu, N., and Krumlauf, R., 1989. The murine and *Drosophila* homeobox gene complexes have common features of organisation and expression. Cell, **57**, 367-378.

Grenier, J. K., Garber, T. L., Warren, R., Whitington, P. M., and Caroll, S., 1997. Evolution of the entire arthropod *Hox* gene set predated the origin and radiation of the onychophoran/arthropod clade. Curr. Biol., **7**, 547-553.

Grobben, C., 1881. Die Entwicklungsgeschichtw von *Cetochilus septentrionalis* Goodsir. Arb. Zool. Inst. Univ. Wien., **3**, 1-40.

Guisande, C., and Harris, R. P., 1995. Effect of total organic content of eggs on hatching success and naupliar survival in the copepod *Calanus helgolandicus*. Limnol. Oceanogr., **40**, 476-482.

Hernández-León, S., Almeida, C., and Montero, I., 1995. The use of aspartate transcarbamylase activity to estimate growth rates in zooplankton. ICES J. Mar. Sci., **52**, 377-383.

Hirche, H. J., 1991. Distribution of dominant copepod species in the Greenland Sea during fall. Polar Biol., **11**, 351-362.

Hirche, H. J., Hagen, W., Mumm, N., and Richter, C., 1994. Mesozooplankton and macrozooplankton distribution and production of dominant herbivorous copepods during spring. Polar Biol., **14**, 491-503.

Holland, P. W., Holland, L. Z., Williams, N. A., and Holland, N. D., 1992. An amphioxus homeobox gene: sequence conservation, spatial expression during development and insights into vertebrate evolution. Dev., **116**, 653-661.

Holland, P. W. H. and Garcia-Fernàndez, J., 1996. *Hox* genes and chordate evolution. Dev. Biol., **173**, 382-395.

Holm-Hansen, O., Sutcliffe, W. H., and Sharp, J., 1968. Measurements of deoxyribonucleic acid in the ocean and its ecological significance. Limnol. Oceanogr., **13**, 507-514.

Huntley, M. E., and Lopez, M. D. G., 1992. Temperature-dependent production of marine copepods: a global synthesis. Am. Nat., **140**, 201-242.

Ianora, A., and Buttino, I., 1990. Seasonal cycles in population abundances and egg production rates in the planktonic copepod *Centropages typicus*. J. Plankton Res., **12**, 473-481.

Ianora, A., Mazzocchi, M. G., and Grottoli, R., 1992. Seasonal fluctuations in fecundity and hatching success in the planktonic copepod *Centropages typicus*. J. Plankton Res., **14**, 1483-1494.

Ianora, A., and Poulet, S., 1993. Egg viability in the copepod *Temora stylifera*. Limnol. Oceanogr., **38**, 1615-1626.

Ianora, A., Poulet, S. A., and Miralto, A., 1995. A comparative study of the inhibitory effect of diatoms on the reproductive biology of the copepod *Temora stylifera*. Mar. Biol., **121**, 533-539.

Ianora, A., Poulet, S. A., Miralto, A., and Grottoli, R., 1996. The diatom *Thalassiosira rotula* affects reproductive success in the copepod *Acartia clausi*. Mar. Biol., **125**, 279-286.

Ianora, A., Miralto, A., and Poulet, S. A., 1999. Are diatoms good or toxic for copepods? Reply to comment Jónasdóttir *et al.* Mar. Ecol. Prog. Ser., **177**, 305-308.

Ingham, 1988. The molecular genetics of embryonic pattern formation in *Drosophila*. Nature, **335**, 25-34.

James, R., Erler, T.,and Kazenwadel, J., 1994. Structure of the Murine Homeobox Gene *cdx-2*. J. Biol. Chem., **269**, 15229-15237.

Jaschnov, V. A., 1970. Distribution of *Calanus* species in the sea of the Northern Hemisphere. International Revue der gesamten Hydrobiol., **55**, 197-212.

Jaschnov, V. A., 1972. On the systematic status of *C. glacialis*, *C. finmarchicus*, and *C. helgolandicus*. Crustaceana, **22**, 279-284.

Jónasdóttir, S. H., 1994. Effects of food quality on the reproductive success of *Acartia tonsa* and *Acartia hudsonica*: laboratory observations. Mar. Biol., **121**, 67-81.

Jónasdóttir, S., and Kiørboe, T., 1996. Copepod recruitment and food composition: do diatoms affect hatching success? Mar. Biol., **125**, 743-750.

Jónasdóttir, S., Kiørboe, T., Tang, K. W., St.John, M., Visser, A. W., Saiz, E., and Dam, H. G., 1998. Role of diatoms in copepod production; good, harmless or toxic? Mar. Ecol. Prog. Ser., **172**, 305-308.

Kania, M. A., Bonner, A. S., Duffy, J. B., and Gergen, J. P., 1990. The *Drosophila* segmentation gene *runt* encodes a novel nuclear regulatory protein that is also expressed in the developing nervous system. Gene and Dev., **4**, 1701-1713.

Kappen, C., Schughart, K., and Ruddle, F. H., 1989. Two steps in the evolution of *Antennapedia*-class vertebrate homeobox genes. Proc. Natl. Acad. Sci. USA, **86**, 5459-5463.

Katsuyama, Y., Sato, Y., Wada, S., and Saiga, H., 1999. Ascidian tail formation requires *caudal* function. Dev. Biol., **213**, 257-268.

Keen, J. N., Pappin, D. J. C., and Evans, L. V., 1988. Amino-acid sequence-analysis of the small subunit of ribulose bisphosphate carboxylase from *fucus* (phaeophyceae). J. Phycol., **24**, 324-327

Kern, M. J., Witte, D. P., Valerius, M. T., Aronow, B. J., and Potter, S. S., 1992. A novel murine homeobox gene isolated by a tissue specific PCR cloning strategy. Nucl. Acids Res., **20**, 5189-5195.

Kim, Y., and Nirenberg, M., 1989. *Drosophila* NK-homeobox genes. Proc. Natl. Acad. Sci. USA., **86**, 7716-7720.

Kimmerer, W. J., 1987. The theory of secondary production calculations for continuously reproducing populations. Limnol. Oceanogr., **32**, 1-13.

Kiørboe, T., Møhlenberg, F., and Hamburger, K., 1985. Bioenergetics of the planktonic copepod *Acartia tonsa*: relation between feeding, egg production and respiration, and composition of specific dynamic action. Mar. Ecol. Prog. Ser., **26**, 85-97.

Kiørboe, T., and Johansen., 1986. Studies of a larval herring (*clupea harengus* L.) patch in the Buchan area. IV. Zooplankton distribution and productivity in relation to hydrographic features. Dana, **6**, 37-51.

Kiørboe, T., Møhlenberg, F., and Tiselius, P., 1988. Propagation of planktonic copepods: production and mortality of eggs. Hydrobiologia, **167/168**, 219-225.

Kleppel, G., S., 1993. On the diets of calanoid copepods. Mar. Ecol. Prog. Ser., **99**, 183-195.

Kocher, T. D., Lee, W.-J., Sobolewska, H., Penman, D., and McAndrew, B., 1998. A genetic linkage map of a Cichlid fish, the Tilapia (*Oreochromis niloticus*). Genetics, **148**, 1225-1232.

Krumlauf, R., 1992. Evolution in the vertebrate Hox homeobox genes. Bio Essays, **14**, 245-252.

Laabir, M., Poulet, S. A., Ianora, A., Miralto, A., and Cueff, A., 1995. Reproductive responses of *Calanus helgolandicus* II. *In situ* inhibition of embryonic development. Mar. Ecol. Prog. Ser., **129**, 97-105.

Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., Sogin, M. L., and Pace, N. R., 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analysis. Proc. Natl. Acad. Sci., USA, **82**, 6955-6959.

Laughon, A., and Scott, M. P., 1984. Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins. Nature, **310**, 25-31.

Lawrence, P. A., and Morata, G., 1994. Homeobox genes: Their function in *Drosophila* segmentation and pattern formation. Cell, **78**, 181-189.

LeMotte, P. K., Kuroiwa, A., Fessler, L. I., and Gehring, W. J., 1989. The homeotic gene *Sex Combs Reduced* of *Drosophila*: gene structure and embryonic expression. EMBO J., **8**, 219-227.

Lewin, B., 1997. Regulation of transcription, in *Genes VI*, (ed. Lewin, B.), Oxford University Press, New York. 847-883.

Lewin, B., 1997a. Nuclear splicing, in *Genes VI*, (ed. Lewin, B.), Oxford University Press, New York. 885-921.

Lewis, E. B., 1978. A gene complex controlling segmentation in *Drosophila*. Nature, **276**, 565-570.

Lindeque, P. K., Harris, R. P., Jones, B. M., and Smerdon, G. R., 1999. Simple molecular method to distinguish the identity of *Calanus* species (Copepoda:Calanoida) at any developmental stage. Mar. Biol., **133**, 91-96.

Macdonald, P. M., Ingham, P., and Struhl, G., 1986. Isolation, structure, and expression of even-skipped-a 2nd pair-rule gene of *Drosophila* containing a homeobox. Cell, **5**, 721-734.

Manak, J. R., and Scott, M. P., 1994. A class act: conservation of homeodomain protein functions. Dev., **(suppl.)**, 61-71.

Manwell, C., Baker, C. M. A., Ashton, P. A., and Corner, E. D. S., 1967. Biochemical differences between *Calanus finmarchicus* and *C. helgolandicus*. J. Mar. Biol. Ass. UK, **47**, 145-169.

Marshall, S. M., and Orr. A. P., 1972. The biology of a marine copepod, *Calanus finmarchicus* (Gunnerus), 2nd edition. Oliver and Boyd, London.

Martinez, I., Skjerdal, T., Dreyer, B., and Aljanabi, S. M., 1997. Genetic structuring of *Panalus borealis* in the North Atlantic. II. RAPD analysis. ICES Rpt. No. CM 1997/T:24.

Mauchline, J., 1997. The Biology of Calanoid Copepods, (eds. Blaxter, J. H. S., Southward, A. J., and Tyler, P. A.). Academic Press, London. 14-48.

McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A., and Gehring, W. G., 1984. A conserved DNA sequence in homeotic genes of the *Drosophila Antennapedia* and bithorax complexes. Nature, **308**, 428-433.

McGinnis, W., and Krumlauf, R., 1992. Homeobox genes and axial patterning. Cell, **63**, 969-976.

McKinnon, A. D., and Thorrold, S. R., 1993. Zooplankton community structure and copepod egg production in coastal waters of the central Great Barrier Reef Lagoon. J. Plankton Res., **15**, 1387-1411.

McLaren, I. A., Sevigny, J.-M., and Corkett, C. J., 1988. Body sizes, development rates, and genome sizes among *Calanus* species. Hydrobiologia, **167/168**, 275-284.

Meyer, A., 1998. *Hox* gene variation and evolution. Nature, **391**, 225-228.

Miralto, A., Ianora, A., and Poulet, S. A., 1995. Food type induces different reproductive responses in the copepod *Centropages typicus*. J. Plankton Res., **17**, 1521-1534.

Miralto, A., Ianora, A., Guglielmo, L., Zagami, G., and Buttino, I., 1998. Egg production and hatching success in the peri-Antarctic copepod *Calanus simillimus*. J. Plankton Res., **20**, 2369-2378.

Miralto, A., Barone, G., Romano, G., Poulet, S. A., Ianora, A., Russo, G. L., Buttino, I., Mazzarella, G., Laabir, M., Cabrini, M., and Giacobbe, M. G., 1999. The insidious effect of diatoms on copepod reproduction. Nature, **402**, 173-176.

Mlodzik, M., and Gehring, W. J., 1987. Expression of the *caudal* gene in the germ line of *Drosophila*: formation of an RNA and protein gradient during early embryogenesis. Cell, **48**, 465-478.

Mlodzik, M., Fjose, A., and Gehring, W. J., 1988. Molecular structure and spatial expression of a homeobox gene from the labial region of the *Antennapedia*-complex. EMBO J., **7**, 2569-2578.

Mouchel-Vielh, E., Rigolot, C., Gibert, J.-M., and Deutsch, J. S., 1998. Molecules and the body plan: the *Hox* genes of Cirripedes (Crustacea). Mol. Phylogen. Evol., **9**, 382-389.

Nüsslein-Volhard, C., 1991. Determination of the embryonic axes of *Drosophila*. Dev., **(suppl. 1)**,1 *et seq.*

Ochman, H., Gerber, A. S., and Hartl, D. L., 1988. Genetics applications of an Inverse Polymerase Chain-Reaction. Genetics, **120**, 621-623.

Orlova, E. L., Ushakov, N. G., Nesterova, V. N., and Pushchaeve, T. Y., 1999. Year-to-year variations in consumption of *C. finmarchicus* by capelin *Mallotus villosus* and peculiarities of the capelin growth in high latitudes in the Barents Sea. Abstract, ICES symposium: Population dynamics of *Calanus* in the North Atlantic – results from the trans-Atlantic study of *Calanus finmarchicus*.

Ota, A. Y., and Landry, M. R., 1984. Nucleic acids as growth rate indicators for early developmental stages of *Calanus pacificus* Brodsky. J. Exp. Mar. Biol. and Ecol., **80**, 147-160.

Palumbi, S. R., and Benzie, J., 1991. Large mitochondrial DNA differences between morphologically similar Penaeid shrimp. Mol. Mar. Biol. Biotech., **1**, 27-34.

Peterson, W. T., 1985. Abundance, age structure and *in situ* egg production rates of the copepod *Temora longicornis* in Long Island Sound, New York. Bull. Mar. Sci., **37**, 726-738.

Pignoni, F., Baldarelli, R. M., Steingrimsson, E., Diaz, R. J., Patapoutian, A., Merriam, J. R., and Lengyel, J. A., 1990. The *Drosophila* gene *tailless* is expressed at the embryonic termini and is a member of the steroid receptor superfamily. Cell, **62**, 151-163.

Planque, B., and Fromentin, J.-M., 1996. *Calanus* and environment in the eastern North Atlantic. I. Spatial and temporal patterns of *C. finmarchicus* and *C. helgolandicus*. Mar. Ecol. Prog. Ser., **134**, 101-109.

Planque, B., Hays, G. C., Ibanez, F., and Gamble, G. C., 1997. Large scale spatial variations in the seasonal abundance of *Calanus finmarchicus*. Deep-Sea Res. I, **44**, 315-326.

Pond, D., Harris, R., Head, R., and Harbour, D., 1996. Environmental and nutritional factors determining seasonal variability in the fecundity and egg viability of *Calanus helogolandicus* in coastal waters off Plymouth, UK. Mar. Ecol. Prog. Ser., **143**, 45-63.

Poulet, S. A., Ianora, A., Miralto, A., and Meijer, L., 1994. Do diatoms arrest embryonic development in copepods? Mar. Ecol. Prog. Ser., **111**, 79-86.

Poulet, S. A., Ianora, A., and Miralto, A., 1995. Reproductive response of *Calanus helgolandicus*. I. Abnormal embryonic and naupliar development. Mar. Ecol. Prog. Ser., **129**, 85-95.

Pouwels, P. H., 1995. Survey of cloning vectors for *Escherichia coli*, in *Essential molecular biology, volume1* (ed. Brown, T. A.). IRL press at Oxford, University press, Oxford. 179-239.

Pownall, M. E., Tucker, A. S., Slack, J. M. W., and Isaacs, H. V., 1996. *EFGF, Xcad3* and *Hox* genes form a molecular pathway that establishes the anteroposterior axis in *Xenopus*. Dev., **122**, 3881-3892.

Raymont, J. E. G., 1983. Plankton productivity in the oceans, volume 2, Zooplankton. Second edition. Pergamon Press. Oxford, 1-24.

Rehnstam, A.-S., Norqvist, A., Wolf-Watz, H., and Hagström, A., 1989. Identification of *Vibrio anguillarum* in fish by using partial 16S rRNA sequences and a specific 16S rRNA oligonucleotide probe. App. Environ. Micro., **55**, 1907-1910.

Reid, L., 1990. From gradients to axes, from morphogenesis to differentiation. Cell, **63**, 875-882.

Rex, M. and Scotting, P. J., 1994. Chick *HoxB3*: deduced amino-acid sequence and embryonic gene expression. Gene, **149**, 381-382.

Rosa, R., Grenier, J. K., Andreeva, T., Cook, C. E., Adoutte, A., Akam, M., Carroll, S. B., Balavoine, G., 1999. Hox genes in branchiopods and priapulids and protosome evolution. Nature, **399**, 772-776.

Runge, J. A., 1985. Should we expect a relationship between primary production and fisheries? The role of copepod dynamics as a filter of trophic variability. Hydrobiologia, **167/168**, 61-71.

Runge, J. A., 1988. Should we expect a relationship between primary production and fisheries? The role of copepod dynamics as a filter of trophic variability. Hydrobiol., **167/168**, 61-71.

Runge, J. A., and Lafontaine, Y. D., 1996. Characterisation of the pelagic ecosystem in surface waters of the northern Gulf of St. Lawrence in early summer: the larval redfish – *Calanus* – microplankton interaction. Fish. Oceanogr., **5**, 1-20.

Runge, J. A., and Roff, J. C., 1999. The measurement of growth and reproductive rates, in *ICES Zooplankton Methodology Manual* (eds. R. Harris, P. H. Wiebe, J. Lenz, H.-R. Skjoldal, M. Huntley), Academic Press.

Rushlow, C. A., Hogan, A., Pinchin, S. M., Howe, K. M., Lardelli, M., and Ish-Horowicz, D., 1989. The *Drosophila hairy* protein acts in both segmentation and bristle patterning and shows homology to N-*myc*. EMBO J., **8**, 3095-3103.

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., and Arnheim, N., 1985. Enzymatic amplification of β-globulin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science, **230**, 1350-1354.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A., 1988. Primer-directed amplification of DNA with a thermostable DNA polymerase. Science, **239**, 487-494.

Saiz, E., Calbert, A., Fara, A., and Berdalet, E., 1998. RNA content of copepods as a tool for determining adult growth rates in the field. Limnol. Oceanogr., **43**, 465-470.

Sambrook, J., Fritsch, E. F., and Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual. 2nd edn. Cold Spring Harbour Laboratory, Cold Spring harbour, New York.

Sanger, F., Nicklen, S., Coulson, A. R., 1977. DNA sequencing with chain-termination inhibitors. Proc. Natl. Acad. Sci. USA, **75**, 5463-5467.

Sapienza, C., and Mague, T. H., 1979. DNA polymerase activity and growth rate in *Artemia saline* nauplii. Limnol. and Oceanogr., **24**, 572-576.

Schierwater, B., Murtha, M., Dock, M., Ruddle, F. H., and Buss, L. W., 1991. Homeoboxes in Cnidarians. J. Exp. Zoo., **260**, 413-416.

Schubert, F. R., Nieselt-Struwe, K., and Gruss, P., 1993. The *Antennapedia*-type homoebox genes have evolved from three precursors separated early in metazoan evolution. Proc. Natl. Acad. Sci. USA, **90**, 143-147.

Schuh, R., Aicher, W., Gaul U., Côté, S., Preiss, A., Maier, D., Seifert, E., Nauber, U., Schröder C., Kemier, R., and Jäckie, H., 1986. A conserved family of nuclear proteins

containing structural elements of the finger protein encoded by *Kruppel*, a *Drosophila* segmentation gene. Cell, **47**, 1025-1032.

Schulz, C., Schröder, R., Hausdorf, B., Wolff, C., and Tautz, D., 1998. A *caudal* homologue in the short germ band beetle *Tribolium* shows similarities to both, the *Drosophila* and the vertebrate *caudal* expression patterns. Dev. Genes Evol., **208**, 283-289.

Scott, M. P., and Weiner, A. J., 1984. Structural relationships among genes that control development: sequence homology between the *Antennapedia*, *Ultrabithorax*, and *fushi tarazu* loci of *Drosophila*. Proc. Natl. Acad. Sci., USA, **81**, 4115-4119.

Scott, M. P., 1994. Intimations of a creature. Cell, **79**, 1121-1124.

Scotting, P. J., Hewitt, M., and Keynes, R. J., 1990. Isolation and analysis of chick homeobox cDNA clones. Nucl. Acids Res., **18**, 3999.

Scrope-Howe, S., and Jones, D. A., 1985. Biological studies in the vicinity of a shallow-sea tidal mixing front v. composition, abundance and distribution of zooplankton in the Western Irish Sea, April 1980 to November 1982. Philosophical transactions of the Royal Society London, **B310**, 501-519.

Seed, B., Parker, R. C., Davidson, N., 1982. Representation of DNA sequences in recombinant DNA libraries prepared by restriction enzyme partial digestion. Gene, **19**, 201-209.

Seimiya, M., Ishiguru, H., Miura, K., Watanabe, Y., and Kurosawa, Y., 1994. Homeobox-containing genes in the most primitive metazoa, the sponges. Eur. J. Biochem., **221**, 219-225.

Sekiguchi, H., McLaren, I. A., and Corkett, C. J., 1980. Relationships between growth rate and production in the copepod *Acartia clausi hudsonica*. Mar. Biol., **58**, 133-138.

Sevigny, J. –M., and McLaren, I. A., 1988. Protein polymorphisms in six species of the genus *Calanus*. Hydrobiol., **167/168**, 267-274.

Shapiro, M. B., and Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucl. Acids. Res., **15**, 7155-7175.

Silver, J., 1991. Inverse Polymerase Chain Reaction, in *PCR; a practical approach* (ed. McPherson, M. J., Taylor, G. R.), IRL Press, Oxford.

Smerdon, G. R., 1998. Towards the molecular analysis of copepod production, in Molecular approaches to the study of the ocean (ed. Cooksey K. C.), Chapman and Hall, London. 319-328.

Slack, J. M. W., 1991. Embryology-Molecule of the moment. Nature, **349**, 17-18.

Slack, J. M. W., Holland, P. W. H., and Graham, C. F., 1993. The zootype and the phylotypic stage. Nature, **361**, 490-492.

Slack, J. M. W., 1997. The anteroposterior system, in *From egg to embryo, 2nd edition* (eds. Barlow, P. W., Bray, D., Green, P. B., and Slack, J. M. W.), Cambridge University press, Cambridge. 236-260.

Slack, J. M. W., 1997. Determinative events in early development, in *From egg to embryo, 2nd edition* (eds. Barlow, P. W., Bray, D., Green, P. B., and Slack, J. M. W.), Cambridge University press, Cambridge. 236-260.

Southern, M., 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol., **98**, 503-517.

Stearns, D. E., Tester, P. A., and Walker, R. L., 1989. Diel changes in the egg production rate of *Acartia tonsa* (Copepoda:calanoida) and related environmental factors in two estuaries. Mar. Ecol. Prog. Ser., **52**, 7-16.

Suazo, A., and Hall, H., G., 1999. Modification of the AFLP protocol applied to honey bee (*Apis mellifera* L.) DNA. Biotechniques, **26**, 704-709.

Sundt, T. M., and Melle, W., 1998. Atlantic observations of *Calanus marshallae* (Copepoda:Calanoida). Mar. Ecol. Prog. Ser., **166**, 207-210.

Sutcliffe, W. H., Jr., 1965. Growth estimates from ribonucleic acid content in some small organisms. Limnol. Oceanogr., **(suppl. 10)**, 253-258.

Takagaki, Y., and Manley, J. L., 1994. A polyadenylation factor subunit is the human homologue of the *Drosophila* suppressor of *forked* protein. Nature, **372**, 471-474.

Tiselius, P., Nielsen, T. G., Breuel, G., Jaanus, A., Korshenko, A., and Witek, Z., 1990. Copepod egg production in the Skagerrak during SKAGEX, May-June. Mar. Biol., **111**, 445-453.

Todd, C. D., Walker, A. M., Wolff, K., Northcott, S. J., Walker, A. F., Ritchie, M. G., Hoskins, R., Abbott, R. J., and Hazon, N., 1997. Genetic differentiation of populations of the copepod sea louse *Lepeophtheirus salmonis* (Køyer) ectoparasitic on wild and farmed salmonids around the coast of Scotland: Evidence from RAPD markers. J. Exp. Mar. Biol. Ecol., **210**, 251-274.

Triantahyllidis, G. V., Criel, G. R. J., Abatzopoulis, T. J., Thomas, K. M., Beardmore, J. A., and Sorgeloos, P., 1997. International study on *Artemia*. LVII. Morphological and molecular characters suggest conspecificity of all bisexual European and North African *Artemia* populations. Mar. Biol., **129**, 477-487.

Triglia, T., Peterson, M. G., and Kemp, D. J., 1988. A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences. Nucl. Acids Res., **16**, 8186.

Trujillo-Ortiz, A. 1995. Alternative method for the calculation of mean time for the assessment of secondary production by true cohort analysis. J. Plankton Res., **17**, 2175-2190.

Uye, S.-I., 1996. Induction of reproductive failure in the planktonic copepod *Calanus pacificus* by diatoms. Mar. Ecol. Prog. Ser., **133**, 89-97.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M., 1995. AFLP: a new technique for DNA fingerprinting. Nucl. Acids Res., **23**, 4407-4414.

Wagner, M. M., Campbell, R. G., and Durbin, E. G., 1999. RNA/DNA ratios of *Calanus finmarchicus* in the Georges Bank region: relationships with *in situ* cholorophyll concentration, egg production, and growth rates. Abstract, ICES symposium: Population dynamics of *Calanus* in the North Atlantic – results from the trans-Atlantic study of *Calanus finmarchicus*.

Ward, D. M., Weller, R., and Bateson, M. B., 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature, **345**, 63-65.

Welsh, J. and M., McClelland, 1990. Fingerprinting genomes using PCR with arbitrary primers. Nuc. Acids Res., **18**, 7213-7218.

Wieschaus, E., 1996. Embryonic transcription and the control of developmental pathways. Genetis, **142**, 5-10.

Williams, R. and Conway, D. V. P., 1980. Vertical distribution *of Calanus finmarchicus* and *C. helgolandicus* (Crustacea: Copepoda). Mar. Biol., **69**, 57-61.

Williams, J. G. K., Kubelik A. R., Livak., K. J., Rafalski, J. A., and Tingey, S. V., 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl. Acids Res., **18**, 6531-6535.

Woese, 1987. Bacterial Evolution. Microbiol. Rev., **51**, 221-271.

Wormington, M., 1993. Poly(A) and translation : development control. Curr. Opinion in Cell Biol., **5**, 950-954.

Xu, X., Xu, P.-X., Suzuki, Y., 1994. A maternal homeobox gene, *Bombyx caudal*, forms both mRNA and protein concentration gradients spanning anteroposterior axis during gastrulation. Dev., **120**, 277-285.

Marine Biology (1999) 133: 91–96
© Springer-Verlag 1999

P. K. Lindeque · R. P. Harris · M. B. Jones · G. R. Smerdon

Simple molecular method to distinguish the identity of *Calanus* species (Copepoda: Calanoida) at any developmental stage

Received: 11 May 1998 / Accepted: 5 August 1998

Abstract Diagnostic morphological characteristics of copepods of the genus *Calanus* are restricted largely to minor variations in secondary sex characteristics. This presents a persistent problem in the identification of individuals to species level, especially for immature stages. We have developed a simple molecular technique to distinguish between the North Atlantic *Calanus* species (*C. helgolandicus*, *C. finmarchicus*, *C. glacialis* and *C. hyperboreus*) at any life stage. Using the polymerase chain-reaction (PCR), the mitochodrial large subunit (16S) ribosomal RNA (rRNA) gene was amplified from individual copepods preserved in ethanol. Subsequent digestion of the amplified products with the restriction enzymes *Dde*I and *Vsp*I, followed by electrophoretic separation in 2% agarose (Metaphor, FMC Ltd), produced a characteristic pattern for each species. The versatility of the method is demonstrated by the unambiguous identification to species of any life stage, from egg to adult, and of individual body parts.

Introduction

Copepods of the genus *Calanus* form a predominant proportion of the zooplankton biomass in the North Atlantic. They play an important role in marine food webs, both as consumers of primary production and as prey species for the larvae of commercial fish (Runge 1988). Indeed, there are believed to be more individuals of the genus *Calanus* alive at any one time than any other animal (Bucklin et al. 1996). Despite the relatively high abundance and ecological importance of the genus in both coastal and open oceanic planktonic assemblages, systematic relationships amongst the species are still unclear. Until the latter half of this century (Manwell et al. 1967; Frost 1974), it was not certain whether *C. helgolandicus* and *C. finmarchicus* were distinct species. *Calanus* species are reproductively isolated, but the manner in which this is attained results in extreme morphological similarity (Fleminger and Hulsemann 1977; Bucklin et al. 1995). Diagnostic morphological characteristics of *Calanus* species are restricted essentially to minor variations in their secondary sex characteristics, presenting a persistent problem in the identification of individuals to species level. Immature *Calanus* are the most problematic and, historically, the geographic location of collection has, at least partially, been relied upon as an indicator of specific identity.

When morphological characters are not practical for routine identification to species level, genetic characters can be used to provide unambiguous taxonomic discrimination. A reliable routine method to identify species is important for establishing their horizontal and vertical ranges (Fleminger and Hulsemann 1977). Such a method would also enable questions about the distribution and abundance of developmental stages to be answered.

Despite morphological similarity, *Calanus* species exhibit considerable base-sequence divergence in the mitochondrial large subunit (16S) ribosomal RNA (rRNA) gene (Bucklin et al. 1995). The 16S rRNA gene has been used to determine phylogenetic relationships in a wide range of organisms (Lane et al. 1985; DeLong et al. 1989; Rehnstam et al. 1989; Ward et al. 1990; Britschgi and Giovannoni 1991), and to examine the molecular systematics of *Calanus* (Bucklin et al. 1992).

Communicated by J.P. Thorpe, Port Erin

P.K. Lindeque (✉) · R.P. Harris · G.R. Smerdon
Centre for Coastal and Marine Sciences,
Plymouth Marine Laboratory,
Prospect Place,
Plymouth, PL1 3DH, Devon, England
Fax: +44 (0) 1752 633 101
e-mail: pkw@wpo.nerc.ac.uk

P.K. Lindeque · M.B. Jones
Plymouth Environmental Research Centre,
Department of Biological Sciences, University of Plymouth,
Plymouth, Drake Circus, Plymouth PL4 8AA, Devon, England

Although this mitochondrial gene varies sufficiently to discriminate closely related species, its intraspecific variation is low enough to warrant its use for species' identification. In this study, we build on the work of Bucklin et al. (1992, 1995), in which the DNA sequence of 16S rRNA genes was determined and used to compare phylogenetic relationships between *Calanus* species. We have further developed this system to produce a method whereby simple molecular techniques can be used to determine unambiguously the species of adults, juveniles and eggs of the dominant *Calanus* species found in the North Atlantic Ocean. Additionally, this technique is useful for the identification of small body parts of *Calanus* species which would enable their identification in gut-content analyses of predators.

To develop our molecular system of *Calanus* identification we used the four species *C. helgolandicus*, *C. finmarchicus*, *C. glacialis* and *C. hyperboreus*. Despite their distinct hydrographic affinities, these species are sympatric in many areas, with overlapping reproductive periods. *C. finmarchicus* is a characteristic northern boreal species that borders *C. glacialis*, primarily an arctic species (Frost 1974). In the eastern North Atlantic, *C. finmarchicus*, *C. glacialis*, and *C. hyperboreus* are sympatric in many areas including the region of the Polar Front. The former two species also co-occur in the region of the Gulf of St. Lawrence, Hudson and Davis Straits (Bucklin et al. 1995). In the eastern North Atlantic, the reproductive range of *C.helgolandicus* overlapped with that of *C. finmarchicus* and *C. glacialis* for appreciable periods in the history of the three species (Fleminger and Hulsemann 1977). *C. helgolandicus*, although a more temperate species, co-occurs extensively with *C. finmarchicus* in many areas of the North Atlantic and in the shelf seas around the United Kingdom (Williams and Conway 1980). The distribution, interactions and different life strategies of these two *Calanus* species are complex: they show converse vertical distributions in areas where they co-exist (Williams and Conway 1980), and their biogeographical boundaries are modified by the North Atlantic Oscillation (Fromentin and Planque 1996). These four species of *Calanus* were chosen for study not only because of their co-occurrence, but also because they constitute a major component of zooplankton in the North Atlantic and North Sea in terms of biomass and trophic role.

Materials and methods

Sample collection and preservation

Individuals of *Calanus helgolandicus*, *C. finmarchicus*, *C. glacialis* and *C. hyperboreus*, collected by net tows, were preserved in absolute ethanol for a minimum of 12 h, with ≃10 adults per 10 ml ethanol. *C. helgolandicus* were collected in April 1996 from a coastal station ≃10 km off Plymouth (western English Channel: 50°15′N; 4°13′W); *C. finmarchicus* were collected in April/May 1996 from Ganavan Bay (Oban, Western Scotland: 56°27′N; 5°27′W); and *C. glacialis* and *C. hyperboreus* were collected in June 1997 from the Norwegian Sea (77°01′N; 33°50′E).

DNA amplification

Amplifications of the 16S rRNA gene were performed on single, preserved individuals without prior purification of the DNA (Bucklin et al. 1995). Individual copepods were removed from ethanol, and rehydrated in 0.5 ml distilled water in a microcentrifuge tube for 6 to 12 h at room temperature. After rehydration, the water was removed and replaced with 34 µl ultrapure water and 5 µl 10x Dynazyme buffer (Flowgen Instruments Ltd.). The copepods were homogenized using a pellet pestle motor (Anachem Ltd.) and incubated at 4 °C overnight. After incubation, the homogenate was transferred to a 0.7 ml tube and the remaining reaction components were added [5 µl 2 m*M* dNTPs (Promega UK Ltd.), 2.5 µl each of primers 16SAR and 16SB2R (100 ng μl^{-1}), and 2 U of Dynazyme (Flowgen Instruments Ltd.)]. The amplification primers used were 16SAR (5′CGCCTGTTTAACAAAAACAT-3′; Palumbi and Benzie 1991) and 16SB2R (5′-ATTCAACATCGAGGT CACAAAC-3′; custom-designed from existing *Calanus* sequences). Amplification was carried out in an Autogene thermocycler (Grant). The cycling parameters included an initial denaturation step of 96 °C (5 min) followed by 40 cycles of 45 °C (2 min), 72 °C (1 min), and 96 °C (1 min). A final annealing phase at 45 °C (2 min) was followed by an extension phase at 72 °C (3 min) and storage at 4 °C until use. 10 µl aliquots of the amplification reactions were analysed by agarose gel electrophoresis (1.5%) to check amplification efficiency.

Sequencing

To determine the most suitable enzyme(s) for restriction-fragment length polymorphism (RFLP) analysis, the amplified 16S rDNA fragments were sequenced for each of the four *Calanus* species. Fragments were cloned into pBluescript SK$^-$, and sequencing was achieved following the chain-termination method (Sanger et al. 1977) using Sequenase II sequencing kit (Amersham).

Restriction mapping

The Genetics Computer Group (GCG) Sequence Analysis Software Package (Devereux et al. 1984) was used to restriction map the sequences of 16S rDNA for *Calanus helgolandicus*, *C. finmarchicus*, *C. glacialis* and *C. hyperboreus*. From these restriction maps, suitable enzymes were selected to differentially digest each of the four species.

Restriction digest

Restriction digests were performed on a 15 µl aliquot from each amplification by the addition of 0.5 µl of 5 *M* NaCl, 2 µl bovine serum albumin (1 mg ml^{-1}, Promega UK Ltd) and 5 U of each restriction enzyme (*Dde*I and *Vsp*I Promega UK Ltd.). Incubations were performed at 37 °C for 1 h. The digestion products were separated by electrophoresis through a 2% Metaphor agarose gel (FMC Ltd., Rockland, Maine, USA) and pre-chilled for 30 min at 4 °C to improve resolution. The gels were observed and photographed on a UV transilluminator.

Results

In this study, the amplification primers 16SAR and 16SB2R, have been used routinely to amplify 16S rDNA from single, preserved *Calanus* individuals. Amplification was successful on whole adults, animals, all copepodite stages, nauplii, eggs and body parts of copepods. Excessive material, resulting from the homogenisation of large adults, although successfully

amplified did in some cases produce less precise banding compared with reactions amplifying smaller amounts of template material.

Nucleotide-base sequences for this amplified region of mitochondrial 16S rDNA were determined for all four *Calanus* spp. (Fig. 1).

The sequenced rDNA regions were mapped for the presence of suitable RFLPs. The chosen restriction enzymes, *Dde*I and *Vsp*I, produced a unique restriction pattern for each *Calanus* species (Fig. 2).

The amplification of 16S rDNA and subsequent RFLP analysis produced a characteristic pattern for each species, providing an unambiguous statement of identity (Fig. 3). The versatility of this molecular method, allowing all life stages of a species including eggs to be identified, is demonstrated in Fig. 4. This method of species identification is also valid for small body parts, including the antennule, leg, thorax and urosome (Fig. 5).

Discussion

A simple molecular technique has been developed involving amplification of 16S rDNA and subsequent RFLP analysis, which can distinguish between *Calanus helgolandicus, C. finmarchicus, C. glacialis and C. hyperboreus* at any life stage. The method is preferable to the RAPD (randomly amplified polymorphic DNA) technique as it does not require prior purification of the DNA to obtain a reproducible result. Such a purification step would preclude the analysis of single eggs and nauplii by limiting amounts of starting material, and would also extend the overall time required for analysis.

Existing 16S rDNA-amplification systems (Palumbi and Benzie 1991; Bucklin et al. 1992) have been modified to provide a cost-effective PCR product suitable for subsequent RFLP analysis. The newly designed 16SB2R primer has a predicted melting temperature (Tm) closer to that of the forward 16SAR primer, allowing the annealing temperature used in the PCR to be increased to 45 °C. This increases the specificity of the amplification product which, in turn, produces a more distinct restriction profile. The reaction volume has been decreased by 50% to reduce the cost of reactants, and the extension phase of thermocycling reduced to save time.

The restriction enzymes were selected for their ability to produce a different restriction profile for each species whilst being optimally active in a single reaction buffer; in this way, simultaneous double-digestion could be performed in a single reaction. Additionally, there is no requirement for purification of the amplified product or buffer exchange prior to restriction digestion. A suitable buffer for activity of the two restriction enzymes is created simply by increasing the salt concentration of the PCR buffer by the addition of NaCl. This system, involving the amplification of rDNA sequence from single individuals and subsequent RFLP analysis, provides an accurate and reliable diagnostic technique for identifying *Calanus* species.

The sequence data obtained for each of the four 16S rDNA fragments amplified differed slightly from previously published sequence data for the same region of 16S rDNA for the same species (Bucklin et al. 1995). Variation was low for *Calanus helgolandicus* compared with *C. helgolandicus* collected previously from the English Channel (1%) and with *C. helgolandicus* collected

Fig. 1 *Calanus* spp. (*chel* = *C. helgolandicus*; *cfin* = *C. finmarchicus*; *cglac* = *C. glacialis*; *chyper* = *C. hyperboreus*). Sequence data for a region of the mitochondrial 16S rRNA gene (. = alignment gap; *n* = any base; **ctnag** = *Dde*I restriction site; **attaat** = *Vsp*I restriction site; primers in *italics*)

```
  cfin16S  cgcctgttta acaaaaacat cgtaaataat atttataata cctgctcagt ..aatattta aacagccgcg ttagtgttaa ggtagcatag
 cglac16S  cgcctgttta acaaaaacat cgtaaataat atttataata cctgctcagt g.aatattta aacagccgcg ttagtgttaa ggtagcatag
  chel16S  cgcctgttta acaaaaacat cgtaaaatag atttatattg cctgctcagt g.aatattta aacagccgca ttagtgttaa ggtagcatag
chyper16S  cgcctgttta acaaaaacat .gtaaattag atttataatg cctgctcaat gaaatattta aatagccgcg ttagtgttaa ggtagcatag

  cfin16S  taattagttt cttaattggg aaataggatg aatggtttca ctaaaatata gtttttatcc tcatttgcga aattttaatc taagtgaaaa
 cglac16S  taattagttt cttaattggg aaatggaatg aatggttt.a ctaaaatatg atatttattc taatttgcga aattttaatc tgagtgaaaa
  chel16S  taattagttt tttaattgga aaatggaatg aatggcccca ctaaagcata gtatttatac taaaaaatga aattttaatt taagtgaaaa
chyper16S  taattagttt tttaattgga aaatggaatg aatggtttca ctaagatatg gtatttatgc taataaatga aattttaatt taagtg.aaa

  cfin16S  tacttagcag ttgtactagg acgagaagac cctatgaagc tggcaaacta ttaat.acat attcctatta tttattagtt tatttttgg
 cglac16S  tactcagaag atatatttag acgagaaga. cctatgaagc tggtagactt ccaatgtaat tatacgatag ttcatgagtt tatttttgg
  chel16S  tacttaaatg atatatttag acgagaagac cctatgaagc tggtagacca taagtgtaat tatttcatag tag.caggtc tatttttgg
chyper16S  tacttaaaag atcttttaag acgagaagac cctatgaagc t.atagacta taaatataat tattataaat ttt.taagtt tatttttgg

  cfin16S  ggtaaaattt aataatacta ttaacacaat tgtactaaat tacatccttt aggaattatg aagaagctcc tctagggata ac.agcatta
 cglac16S  ggtaaaattt aataatagta ttaatattgg cttactaaat aatatcctct tggaattatg aaaaagctcc tctagggata acaagcatta
  chel16S  ggtaaaattt aataattata ttaatacaga tttgttcaaa cttatccttc tggaattatg aataagctcc tctagggata ac.agcatta
chyper16S  ggtaaaattt aataatttta tttaaataag cttattttaa ttgatccttt aggaattatg aaaaagctcc tctagggata ac.agcatta

  cfin16S  tgcttaaaag agttcttatc agaataagcg tttgtgacct cgatgttgaa t
 cglac16S  tgcttaaaag agttcttatc agaataagcg tttgtgacct cgatgttgaa t
  chel16S  tgcttaaaag agttcttatc agaataagcg tttgtgacct cgatgttgaa t
chyper16S  tacctataag agttcttatc agaataggtg tttgtgacct cgatgttgaa t
```

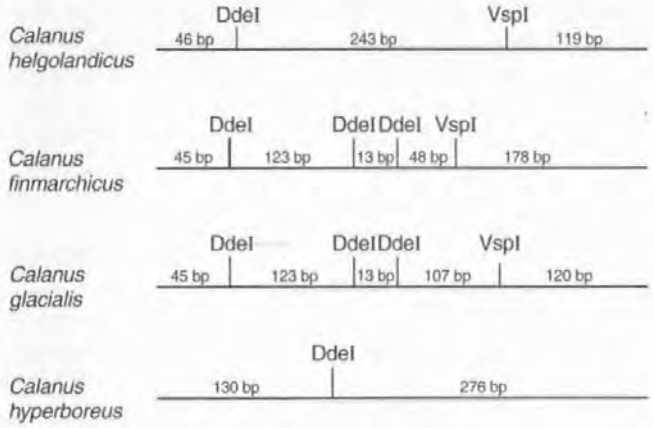


Fig. 2 *Calanus* spp. Restriction map of a region of mitochondrial 16S rDNA for *C. helgolandicus*, *C. finmarchicus*, *C. glacialis*, and *C. hyperboreus*. Both *Dde*I and *Vsp*I restriction sites are indicated. Resulting fragments are shown in base pairs (*bp*)

from the Azores (1.7%). Our sequence data for *C. finmarchicus* also varied only slightly from that for *C. finmarchicus* collected from the Gulf of Maine (0.03%) and from the Gulf of St. Lawrence (2.3%). *C. hyperboreus* showed a small variation of 2.3% from *C. hyperboreus* collected in the Gulf of St. Lawrence. *C. glacialis* displayed a slightly greater variation (9.9%) compared with the same species collected from the Gulf of St. Lawrence.

The differences are small in the case of *Calanus helgolandicus*, *C. finmarchicus* and *C. hyperboreus*, but significant in the case of *C. glacialis*. The latter difference may reflect the site of collection: the original sequence data were obtained for *C. glacialis* from the Gulf of St. Lawrence and our sequence data being obtained from *C. glacialis* collected from the Norwegian Sea. Although this difference is not as significant at that recorded between species (Bucklin and Lajeunesse 1994), it does pose

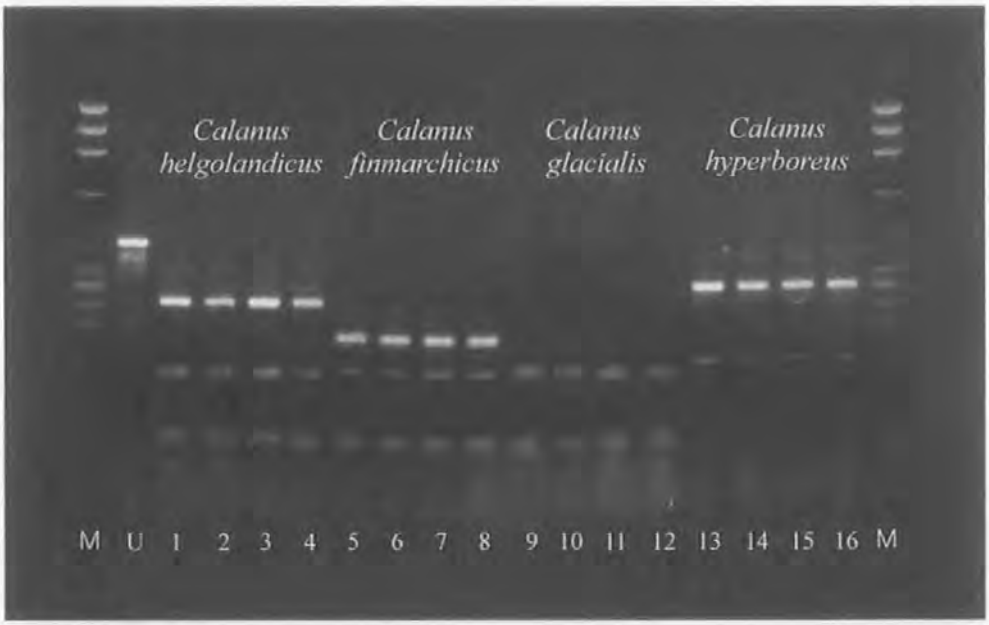


Fig. 3 *Calanus* spp. Amplification of 16S rDNA from whole adults and subsequent RFLP analysis with restriction endonucleases *Dde*I and *Vsp*I (*Lanes 1 to 4, C. helgolandicus; Lanes 5 to 8, C. finmarchicus; Lanes 9 to 12, C. glacialis; Lanes 13 to 16, C. hyperboreus; M*, øX174/*Hae*III DNA size-markers; *U*, amplified uncut *C. helgolandicus* DNA)

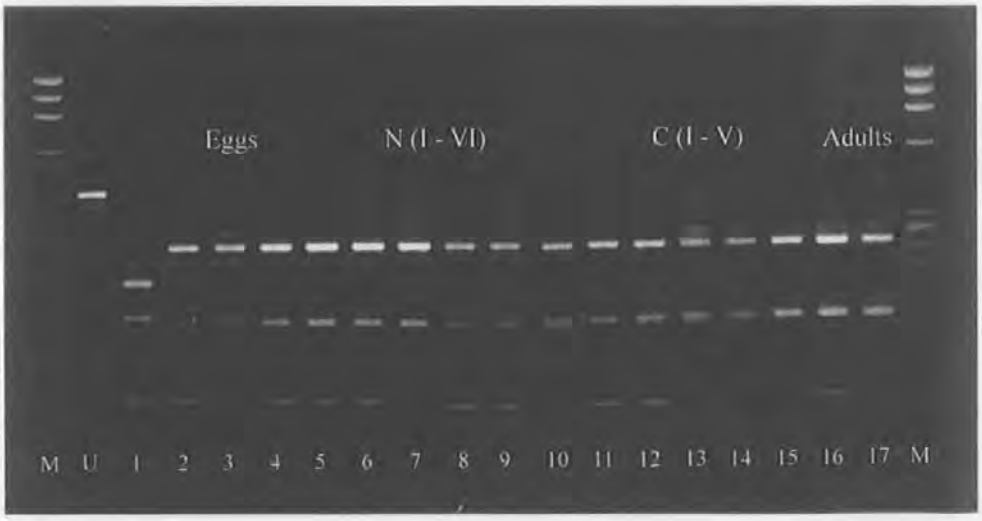


Fig. 4 *Calanus* spp. Amplification of 16S rDNA from major developmental stages of *C. helgolandicus* and subsequent RFLP analysis with *Dde*I and *Vsp*I. Restriction pattern produced by *C. finmarchicus* is shown for comparison (*Lane 1*) (*Lanes 2 to 4* individual eggs; *Lanes 5 to 10* Nauplius Stages I to VI, respectively; *Lanes 11 to 15* Copepodite Stages I to V respectively; *Lanes 16 to 17* Adult copepods; *M*, øX174/*Hae*III DNA size-markers; *U*, amplified uncut *C. helgolandicus* DNA)

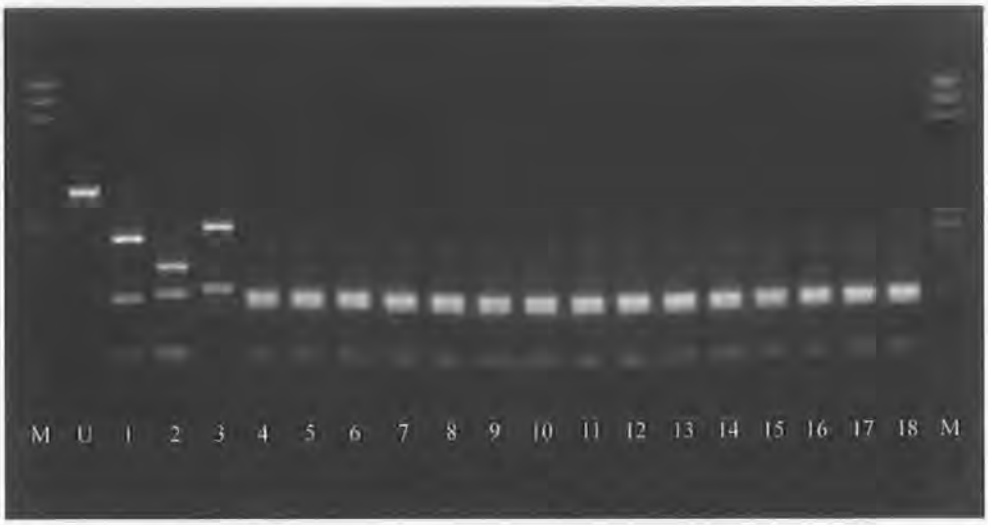


Fig. 5 *Calanus* spp. Amplification of 16 S rDNA from body parts of *C. glacialis* and subsequent RFLP analysis using *Dde*I and *Vsp*I. Restriction pattern produced by *C. helgolandicus* (*Lane 1*), *C. finmarchicus* (*Lane 2*) and *C. hyperboreus* (*Lane 3*) are shown for comparison (*Lanes 4 to 6* individual antennules; *Lanes 7 to 9* individual legs; *Lanes 10 to 12* individual urosomes; *Lanes 13 to 15* individual thoraxes; *Lanes 16 to 18* individual eggs; *M*, øX174/*Hae*III DNA size-markers; *U*, amplified uncut *C. helgolandicus* DNA)

some interesting questions about the phylogenetic relationships of geographically diverse *Calanus* species.

There are many practical applications for this simple molecular technique. It has been well demonstrated that environmental factors such as temperature, oceanic currents, turbulence, wind stress and nutrients can considerably influence the population composition of zooplankton (Colebrook 1982, 1985; Dickson et al. 1988; Fromentin and Planque 1996). Different combinations of *Calanus* species are known to co-occur (Frost 1974; Fleminger and Hulsemann 1977; Williams and Conway 1980). An unambiguous method to distinguish between *Calanus* species at any life stage will increase our understanding of the effect of these environmental factors on zooplankton assemblages, and thus enable a better understanding of their distribution, vertical migration and exploitation of habitat. This in turn will permit further investigations into how environmental changes affect individual zooplankton species with different physiological and biological properties.

Copepods, particularly eggs and nauplii, play an important role in marine food webs by providing an important food source for many animals including fish larvae, small species of fishes and euphausiids. Many experiments involve investigation of the gut content of such animals to determine their diet and role in the ecosystem. *Calanus* spp. eggs can remain undigested during the gut passage (Conway et al. 1994) and can, therefore, be identified easily using RFLP. In predators such as euphausiids, whose food is well masticated, only small body parts of *Calanus* spp. are found in the gut. In such cases, this molecular method, enabling identification of species from fragmentary material, will be particularly valuable.

In conclusion, our method for the identification of the North Atlantic *Calanus* species is reliable, reproducible and relatively simple and inexpensive. It has the benefit that analysis is performed on individuals preserved in ethanol, making it suitable for use on *Calanus* samples collected at sea.

Acknowledgements This work was funded by the NERC special topic PRIME (Plankton Reactivity in the Marine Environment), and was partially supported by funding from the European Commission through the TASC project (MAS3-CT95-0039). We are grateful to Dr. H.J. Hirche for supplying samples *of Calanus glacialis* and *C. hyperboreus* from the Norwegian sea. We would also like to thank D. Conway for his help in initially confirming the identity of the *Calanus* species. This work benefitted from the use of the SEQNET facility.

References

Britschgi TB, Giovannoni SJ (1991) Phylogenetic analysis of a natural marine bacterioplankton population by rRNA gene cloning and sequencing. Appl envirl Microbiol 57: 1707–1713

Bucklin A, Frost BW, Kocher TD (1992) DNA sequence variation of the mitochondrial 16S rRNA in *Calanus* (Copepoda; Calanoida): intraspecific and interspecific patterns. Molec mar Biol Biotechnol 1: 397–407

Bucklin A, Frost BW, Kocher TD (1995) Molecular systematics of six *Calanus* and three *Metridia* species (Calanoida: Copepoda). Mar Biol 121: 655–664

Bucklin A, Lajeunesse TC (1994) Molecular genetic variation of *Calanus pacificus* (Copepoda; Calanoida): preliminary evaluation of genetic structure and sub-specific differentiation based on mtDNA sequences. Rep Calif coop ocean Fish Invest 35: 45–51

Bucklin A, Sundt RC, Dahle G (1996) The population genetics of *Calanus finmarchicus* in the North Atlantic. Ophelia 44: 29–45

Colebrook JM (1982) Continuous plankton records: seasonal variations in the distribution and abundance of plankton in the North Atlantic Ocean and the North Sea. J Plankton Res 4: 435–462

Colebrook JM (1985) Continuous plankton records: overwintering and annual fluctuations in the abundance of zooplankton. Mar Biol 84: 261–265

Conway DVP, McFadzen IRB, Tranter PRG (1994) Digestion of copepod eggs by larval turbot *Scophthalmus maximus* and egg viability following gut passage. Mar Ecol Prog Ser 106: 303–309

DeLong EF, Wickham GS, Pace NR (1989) Phylogenetic stains: ribosomal RNA-based probes for the identification of single cells. Science, NY 24: 1360–1363

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence-analysis programs for the VAX. Nucleic Acids Res 12: 387–395

Dickson RR, Kelly PM, Colebrook JM, Wooster WS, Cushing DH (1988) North winds and production in the eastern North Atlantic. J Plankton Res 10: 151–169

Fleminger A, Hulsemann K (1977) Geographical range and taxonomic divergence in North Atlantic *Calanus* (*C. helgolandicus*, *C. finmarchicus* and *C. glacialis*). Mar Biol 40: 233–248

Fromentin J-M, Planque B (1996) *Calanus* and environment in the eastern North Atlantic. II. Influence of the North Atlantic Oscillation on *C. finmarchicus* and *C. helgolandicus*. Mar Ecol Prog Ser 134: 111–118

Frost BW (1974) *Calanus marshallae*, a new species of calanoid copepod closely allied to the sibling species *C. finmarchicus* and *C. glacialis*. Mar Biol 26: 77–99

Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analysis. Proc natn Acad Sci USA 82: 6955–6959

Manwell C, Baker CMA, Ashton PA, Corner EDS (1967) Biochemical differences between *Calanus finmarchicus* and *C. helgolandicus*. J mar biol Ass UK 47: 145–169

Palumbi SR, Benzie J (1991) Large mitochondrial DNA differences between morphologically similar penaeid shrimp. Molec mar Biol Biotechnol 1: 27–34

Rehnstam A-S, Norqvist A, Wolf-Watz H, Hagström A (1989) Identification of *Vibrio anguillarum* in fish by using partial 16S rRNA sequences and a specific 16S rRNA oligonucleotide probe. Appl envirl Microbiol 55: 1907–1910

Runge JA (1988) Should we expect a relationship between primary production and fisheries? The role of copepod dynamics as a filter of trophic variability. Hydrobiologia 167/168: 61–71

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc natn Acad Sci USA 75: 5463–5467

Ward DM, Weller R, Bateson MB (1990) 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature, Lond 345: 63–65

Williams R, Conway DVP (1980) Vertical distribution *of Calanus finmarchicus* and *C. helgolandicus* (Crustacea: Copepoda). Mar Biol 60: 57–61