Biotechnology and Genetics in Fisheries and Aquaculture

Biotechnology and Genetics in Fisheries and Aquaculture

A.R. Beaumont and K. Hoare

School of Ocean Sciences
University of Wales, Bangor, UK

Blackwell
Science

Editorial Offices:
9600 Garsington Road, Oxford OX4 2DQ
Tel: 01865 776868
Blackwell Publishing, Inc., 350 Main Street,
Malden, MA 02148-5018, USA
Tel: +1 781 388 8250
Iowa State Press, a Blackwell Publishing
Company, 2121 State Avenue, Ames,
Iowa 50014-8300, USA
Tel: +1 515 292 0140
Blackwell Publishing Asia Pty Ltd,
550 Swanston Street, Carlton South,
Victoria 3053, Australia
Tel: +61 (0)3 9347 0300
Blackwell Wissenschafts Verlag,
Kurfürstendamm 57, 10707 Berlin, Germany
Tel: +49 (0)30 32 79 060

First published 2003 by Blackwell Science Ltd

Library of Congress
Cataloging-in-Publication Data
is available

0-632-05515-4

A catalogue record for this title is available from the British Library

Set in Times and produced by
Gray Publishing, Tunbridge Wells, Kent
Printed and bound in Great Britain by
MPG Books, Bodmin, Cornwall

For further information on
Blackwell Science, visit our website:
www.blackwell-science.com

Biotechnology and Genetics in Fisheries and Aquaculture

List of boxes

Preface

The idea for this book was spawned by marine biology graduates at the School of Ocean Sciences, University of Wales Bangor, who proposed that A.R.B.'s Genetics in Aquaculture lecture course notes be packaged into a handbook. What seemed a relatively simple task has, of course, expanded into a larger enterprise. As with all spawnings in aquaculture, there are bound to be some instances of less than perfect development and for this we accept full responsibility. However, we hope that we have produced an introductory-level text which can explain to both students and professionals in fisheries and aquaculture what the new technologies in molecular biology and genetics have to offer.

The authors would like to thank the following for granting permission to use material in this book: Drs Ann Wood, Karen Abey, Halina Sobolewska, Shelagh Malham and Craig Wilding, and Chris Beveridge; Professors John Avise and Steve Karl; copyright holders *The Journal of Shellfish Research*, Cambridge University Press, The American Association for the Advancement of Science, The National Research Council of Canada Research Press, Elsevier Science and The Washington Sea Grant Program, University of Washington.

We are grateful to David Roberts and Geraint Williams of the School of Ocean Sciences, University of Wales Bangor for converting our sketches into publishable illustrations. Finally, we thank Nigel Balmforth of Blackwell Science for his encouragement and patience during the preparation of this book.

A.R. Beaumont & K. Hoare

'How inappropriate to call this planet Earth,
when it is clearly Ocean.'

Arthur C. Clarke

Contents

Chapter 1
What is Genetic Variation?

Have you ever seen someone who looks and sounds exactly like you? Have you ever seen your 'spitting image'? Unless you are one of a pair of monozygotic twins (twins produced by the division of a single egg) you will not have done so. It is commonly accepted that all humans are different – indeed all humans there have ever been were unique and were different from all humans living today. If this is true for *Homo sapiens*, is it also true for other sexually reproducing organisms? The answer is yes. Every salmon (*Salmo salar*) is different from every other salmon that has ever lived. Every mussel (*Mytilus edulis*) is different from every other mussel that has ever lived. This uniqueness of individuals within a species is the consequence of two factors: one is deoxyribose nucleic acid (DNA) and the other is sexual reproduction. These two factors produce and maintain the genetic diversity within a species, and an understanding of this is fundamental to our ability to sustainably exploit species of plants and animals.

Deoxyribose nucleic acid: DNA

The discovery by Watson and Crick of the structure of DNA in 1953 was a landmark in our understanding of how genetic information passes from generation to generation. In the half century since then, the fields of molecular biology and genetics have become inextricably linked and developments, particularly over the past 25 years, have opened up the potential of DNA biotechnology.

The structure of DNA enables it to carry the information for a cell to reproduce itself. It is a polymeric molecule, that is, made up of a chain of subunits, consisting of chains of nucleotide monomers. Each nucleotide contains a base, along with a sugar (deoxyribose) and a phosphate group (Fig. 1.1). There are four individual bases, adenine, guanine, thymine and cytosine and they are usually referred to by their first letter abbreviations, A, G, T and C. Two of the bases, A and G, have a double-ring structure and are known as purines. The other two bases, T and C, are pyrimidines with a single carbon–nitrogen ring.

Each nucleotide is a single unit that joins with neighbouring nucleotides in a linear fashion to make up a polynucleotide chain. Particular carbon atoms in the 5-carbon structure of deoxyribose are referred to by numbers, 1' (one prime) to 5'. The link between nucleotides is formed when the 5' of one bonds to the 3' of the next via a phosphodiester bond (Fig. 1.1). It is the sequence of the four bases in a polynucleotide chain which acts as the code for genetic information.

The complete DNA molecule actually consists of two polynucleotide chains, or *strands*, wrapped around each other in the form of a double helix. The sugar + phosphate backbones are at the outside of the molecule while the bases point towards the

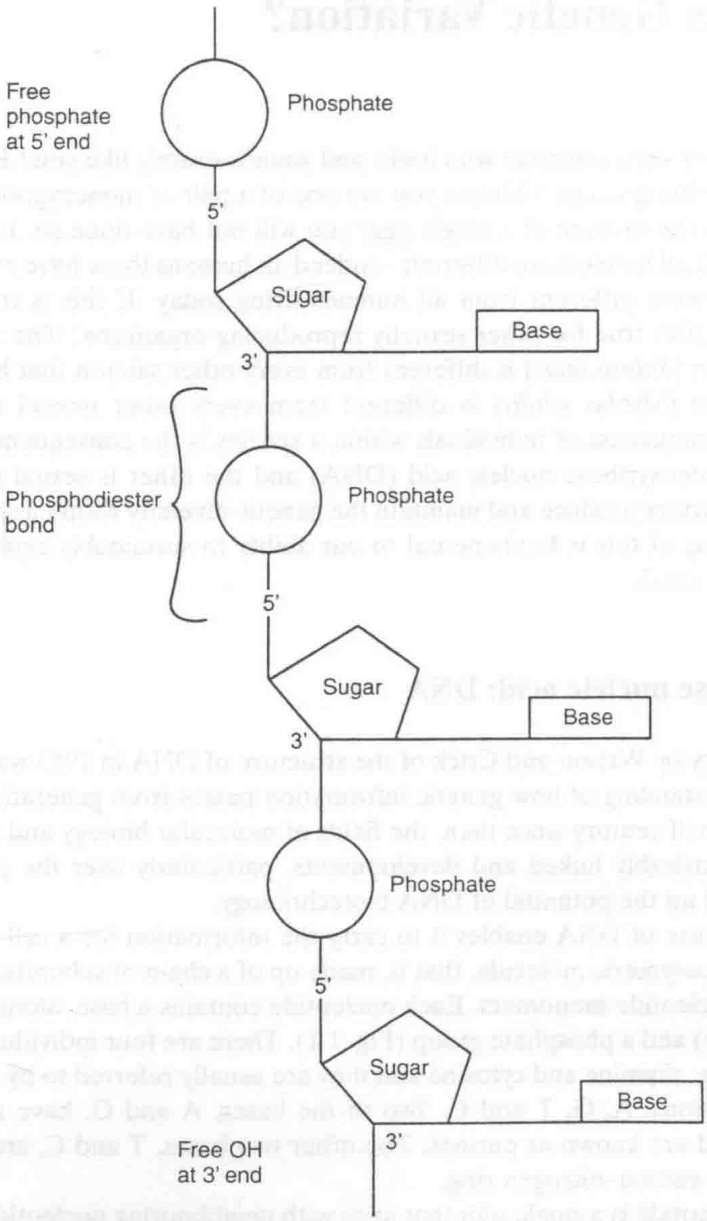

Fig. 1.1 The structure of DNA. Each nucleotide consists of a sugar, a phosphate and a base. Nucleotides are joined by a phosphodiester bond between the 5' of one ribose sugar and the 3' of the next. The chain of nucleotides therefore has a 3' and a 5' end.

middle of the structure; the two strands of the molecule run in opposite directions (Fig. 1.2).

The functional beauty of the DNA molecule is a result of complementary base pairing where G can only bond with C, and A can only bond with T, at the middle of the molecule. It means that the two strands are complementary such that the base

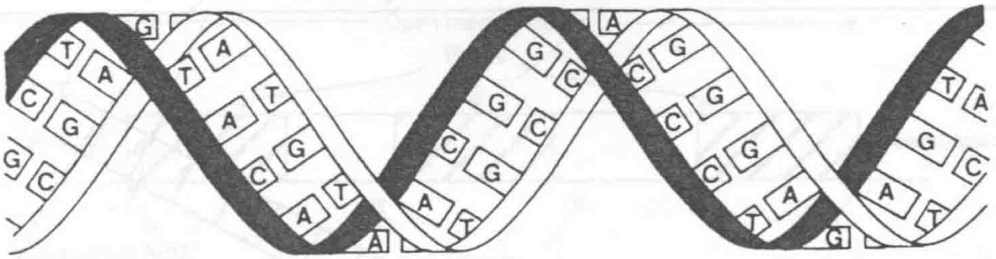

Fig. 1.2 The structure of DNA. Two polynucleotide strands are wrapped around each other in the form of a double helix. Complementary base pairing occurs between the two strands such that guanine (G) always bonds with cytosine (C) and adenine (A) always bonds with thymine (T). (Modified from Utter *et al.* (1987) Interpreting genetic variation detected by electrophoresis. In: *Population Genetics and Fishery Management* (eds. N. Ryman & F. Utter), pp. 21–45, with permission from Washington Sea Grant Program, University of Washington.)

sequence of one strand predicts and determines the base sequence of the other strand. Because one strand predicts the other it can be used to replicate the sequence.

The replication process produces daughter molecules, each of which has one parental strand and one copied strand. This is called semi-conservative replication. Replication of DNA takes place every time a cell divides. The cell's entire DNA is progressively unwound revealing short single-stranded regions which can be copied by DNA polymerase enzymes. Unwinding does not begin at the ends of the molecule, but at points called replication origins, and it then proceeds from these points along the DNA. The new strands of DNA being synthesised during replication are always synthesised in the 5' to 3' direction. This means that as the original strands separate, one new strand can be continuously synthesised against its copy strand (the leading strand) while the other has to be synthesised intermittently in short lengths as enough copy strand (the lagging strand) becomes available (Fig. 1.3).

Considering the enormous numbers of bases and coded information in the DNA of a cell, replication needs to be extremely accurate. Even a very small incidence of mistakes in copying would result in the loss of important genetic information within a few cell divisions. However, during the replication process various proofreading activities take place and almost all errors are corrected by removing the incorrect base and inserting the correct one. In spite of proofreading, a few errors are inevitable when such high numbers of bases are to be copied and it is estimated that about one in every 3 billion bases is incorrectly inserted. Such errors are called point mutations and they can also be induced by certain chemicals and radioactivity. Although there are very few of them, they are nevertheless the fundamental source of variation which fuels the process of evolution. Without such errors, no genetic change at the DNA level would take place, but with too many errors daughter cells would too often be non-viable and the organism carrying that DNA would soon become extinct.

Functional sequences only represent a small fraction of the total genome, for example around 3% in humans. The rest is made up of what has been called 'junk

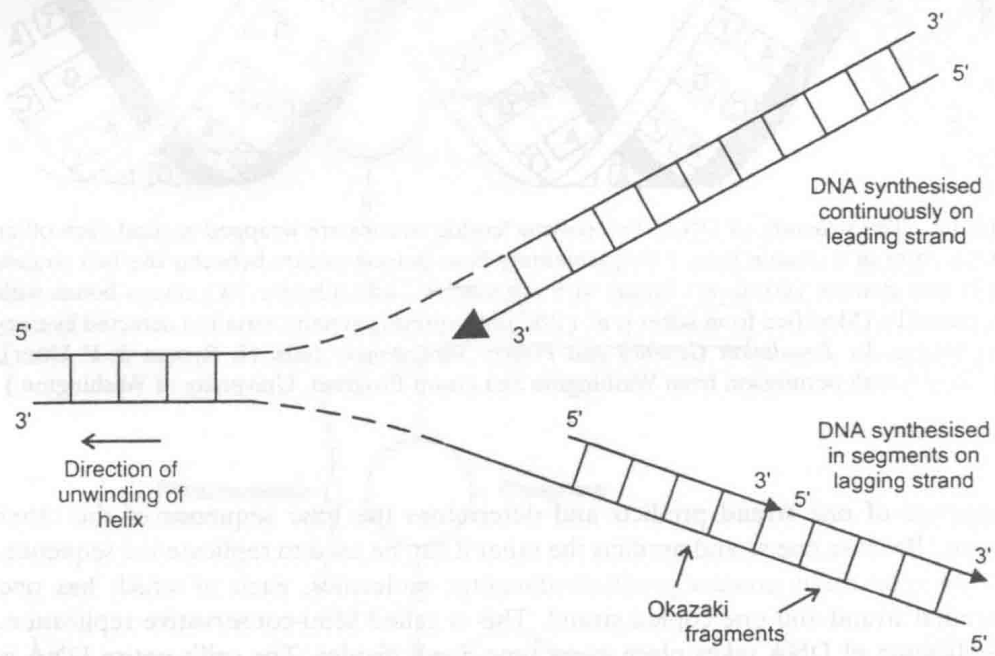

Fig. 1.3 Replication of DNA. As the DNA double helix unwinds, new DNA is synthesised continuously in a 5' to 3' direction on the leading strand and in 5' to 3' directed segments (Okazaki fragments) on the lagging strand.

DNA'. Whether all of it is really 'junk' is not known, but it is possible that much of it will have some, as yet undiscovered, function in the organism. Some of this junk DNA consists of pseudogenes, genes that for some reason or another have become non-functional. Yet other parts of non-coding DNA consist of dispersed or clustered repeated sequences of varying length, from one base pair (bp) to thousands of bases (kilobases, kb) in length. The dispersed repeated sequences occur as copies spread across the genome and can be categorised as long or short interspersed nuclear elements (LINE or SINE), long terminal repeats (LTR) and DNA transposons. The clustered repeated sequences, where the repeated sequence occurs in tandem copies, are classed as satellites, minisatellites or microsatellites depending on the length of the repeat unit, and these have turned out to be useful genetic markers, as will be explained in later chapters. Between them, these repeated elements can constitute up to 40% of the genome.

A gene is a unit of information which is held as a code in a discreet segment of DNA. This code specifies the amino acid sequence of a protein. Scientists were surprised to discover quite early on that the sequence information for a single gene was not continuous along the DNA, but was interspersed with pieces of non-coding sequence. The coding parts of a gene sequence are exons, and the non-coding parts are introns (Fig. 1.4). Before a gene can be expressed, the DNA that encodes it has to be transcribed into RNA.

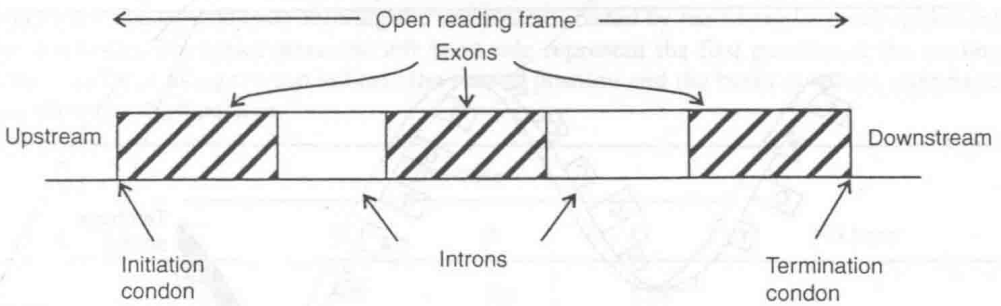

Fig. 1.4 Generalised structure of a gene. The open reading frame (ORF) for a gene begins with an upstream initiation codon and ends with a downstream termination codon. Many genes have a region, or regions, of non-coding DNA within them. These introns are spliced out of the messenger RNA during transcription so that only the codons within the exons are translated into amino acids.

Ribose nucleic acid: RNA

The structure of ribose nucleic acid (RNA) is similar to that of DNA (deoxyribose nucleic acid) except that (a) the sugar is ribose instead of deoxyribose, (b) in the place of thymine, a similarly structured base called uracil (U) is present and (c) the molecule consists of only a single polynucleotide strand. RNA molecules are produced by the process of transcription of the linear sequence of bases in DNA and are then used in the translation of that sequence into a chain of amino acids that go to make up a protein. The type of RNA transcribed from the sequence is called messenger RNA (mRNA) and translation of this sequence into a string of amino acids is undertaken by ribosomal RNA (rRNA) and transfer RNA (tRNA) molecules.

During transcription of the DNA, an RNA copy is made of one of the strands of DNA (Fig. 1.5). The two strands of DNA are called the template strand and the non-template strand. Other names for the non-template strand are the sense (+) strand or the coding strand. The RNA is synthesised by RNA polymerase enzymes using the template strand and is therefore a copy of the non-template (sense or coding) strand of DNA. Because this RNA is a direct copy of the DNA it will contain both the coding (exons) and the non-coding sequences (introns) of the gene. Introns are removed from this pre-messenger RNA and the subsequent molecule is the final mRNA. The mRNA molecules are transported from the nucleus into the cytoplasm where the message is translated into a sequence of amino acids by rRNA in bodies known as ribosomes. Amino acids are brought to the ribosomes by tRNA molecules, each specifying a particular amino acid (Fig. 1.5), and synthesised, in the presence of rRNA, into a linear sequence.

The detailed mechanics and biochemistry of the processes of transcription and translation are outside the scope of this book, but can be found in most standard genetic texts. Some modern texts use the term 'gene expression' to encompass both of these processes and their various controlling steps. For the purposes of this book, the reader need only appreciate the key concept that a sequence of bases in DNA

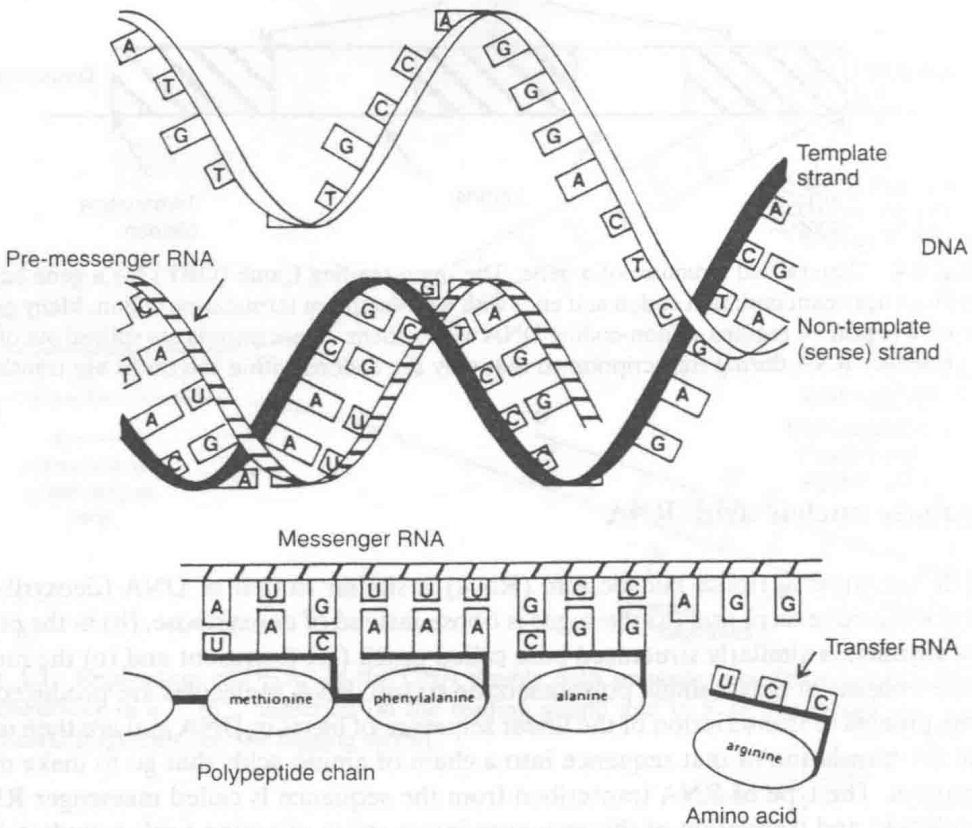

Fig. 1.5 Transcription and translation of DNA. Introns are removed from pre-messenger RNA before translation takes place. The polypeptide chain is formed from amino acids coded for by the messenger RNA and brought together by transfer RNA. (Modified from Utter *et al.* (1987) Interpreting genetic variation detected by electrophoresis. In: *Population Genetics and Fishery Management* (eds N. Ryman & F. Utter), pp. 21–45, with permission from Washington Sea Grant Program, University of Washington.)

leads, by a direct copying process involving RNA, to the production of a sequence of amino acids, the building blocks of proteins. This is what has been called the central dogma: information is transferred from DNA to RNA to protein.

What is the genetic code?

How are the four bases (A, C, G and T) in DNA organised to provide an unambiguous code for the 20 amino acids present in proteins? The 'words' of the code consist of three bases. There are $4^3 = 64$ possible combinations of the four bases into a triplet code and it is these 64 triplet codons which define the 20 amino acids. Because there are more than 20 codons, the genetic code has some redundancy – most amino acids are coded for by more than one codon. The codons are written using the symbol U,

Table 1.1 The genetic code showing the amino acids coded by the 64 triplet combinations of the four bases. The bases down the left hand side represent the first position in the reading frame, the bases along the top indicate the second position and the bases down the right-hand side show the third position

	2nd base				
1st base	U	C	A	G	3rd base
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Abbreviations for amino acids: Alanine (Ala), Arginine (Arg), Asparagine (Asn), Aspartic acid (Asp), Cysteine (Cys), Glutamic acid (Glu), Glutamine (Gln), Glycine (Gly), Histidine (His), Isoleucine (Ile), Leucine (Leu), Lysine (Lys), Methionine (Met), Phenylanaline (Phe), Proline (Pro), Serine (Ser), Threonine (Thr), Tryptophan (Trp), Tyrosine (Tyr), Valine (Val).

for uracil (in mRNA), rather than T, for thymine (in DNA). Three codons (UAA, UAG and UGA) do not encode amino acids but act as signals for protein synthesis to stop and are called termination codons or stop codons. The triplet AUG codes for methionine (formyl methionine in bacteria and mitochondria) and is the signal for protein synthesis to start. It is thus the initiation codon which sets the reading frame. The amino acid sequence of all proteins therefore starts with methionine but this is sometimes removed later. Details of the amino acids encoded by the various codons are given in Table 1.1. Note that the redundancy of the code is not random. In particular, the first two bases of the codons for an amino acid are usually the same. It is generally only the third base which varies.

Protein structure

Proteins have many tasks. Some form the structure of tissues, others – the enzymes – act as extremely specific catalysts of biochemical reactions, and yet other proteins,