Lessons from the large literature on plant reproductive biology have already shown surprising parallels in marine systems, and will be a productive guide to future research.

Acknowledgements

I thank Ed Metz, Vic Vacquier and my parents for discussion about gametes, and Andrew Martin, Craig Moritz, Hampton Carson, Richard Strathmann and Mike Hadfield for comments on the manuscript. Supported by grants from NSF and the Whitehall Foundation.

References

1 Carson, H.L. and Templeton, A.R. (1984) *Annu. Rev. Ecol. Syst.* 15, 97–131
2 Barton, N.H. and Charlesworth, B. (1984) *Annu. Rev. Ecol. Syst.* 15, 133–164
3 Kay, E.A. and Palumbi, S.R. (1987) *Trends Ecol. Evol.* 2, 183–187
4 Otte, D. and Endler, J.A., eds (1989) *Speciation and Its Consequences*, Sinauer Associates
5 Barton, N.H. and Hewitt, G.M. (1989) *Nature* 341, 497–502
6 Cracraft, J. (1986) *Evolution* 40, 977–996
7 Grant, V. (1981) *Plant Speciation* (2nd edn), Columbia University Press
8 Gottlieb, L.D. (1984) *Am. Nat.* 123, 681–709
9 Carson, H.L. (1985) *Syst. Bot.* 10, 380–390
10 Levin, D.A. (1978) *Evol. Biol.* 11, 185–317
11 Waser, N.M. and Price, M.V. (1989) *Evolution* 43, 1097–1109
12 Patterson, H.E.H. (1985) in *Species and Speciation* (Vrba, E.S., ed.), pp. 21–29, Transvaal Museum Monograph No. 4
13 Jablonski, D. (1986) *Bull. Mar. Sci.* 39, 565–587
14 Mayr, E. (1954) *Evolution* 8, 1–18
15 Hilbish, T.J. and Koehn, R.K. (1985) *Evolution* 39, 1302–1317
16 Reeb, C. and Avise, J.C. (1990) *Genetics* 124, 397–406
17 Baker, C.S. *et al.* (1990) *Nature* 344, 238–240
18 Palumbi, S.R. and Wilson, A.C. (1990) *Evolution* 44, 403–415
19 Palumbi, S.R. and Kessing, B.D. (1991) *Evolution* 45, 1790–1805
20 Palumbi, S.R. and Metz, E. (1991) *Mol. Biol. Evol.* 8, 227–238
21 Springer, V. (1988) *Smithson. Contrib. Zool.* 465, 1–134
22 Metz, E.C., Yanagimachi, H. and Palumbi, S.R. in *Proceedings of the 7th International Echinoderm Conference* (Yanigisawa, T. *et al.*, eds), A.A. Balkema Press (in press)
23 Minor, J., Gao, B. and Davidson, E. (1989) in *The Molecular Biology of Fertilization* (Schatten, H. and Schatten, G., eds), pp. 78–88, Academic Press
24 Vacquier, V.D., Corner, K.R. and Stout, C.D. (1990) *Proc. Natl Acad. Sci. USA* 87, 5792–5796
25 Pandey, K.K. (1972) *Theor. Appl. Genet.* 42, 250–261
26 Garbers, D.L. (1989) *Annu. Rev. Biochem.* 58, 719–742
27 Strathmann, R.R. (1985) *Annu. Rev. Ecol. Syst.* 16, 339–361
28 Kirkpatrick, M. and Ryan, M.J. (1991) *Nature* 350, 33–38
29 Day, A.J. and Bayne, B.L. (1988) *Mar. Biol.* 99, 93–100
30 Snyder, T.P. and Gooch, J.L. (1973) *Mar Biol.* 22, 177–182
31 Berger, E.M. (1973) *Biol. Bull.* 145, 83–9
32 Avise, J.C., Reeb, C.A. and Saunders, N. (1987) *Evolution* 41, 991–1002
33 Burton, R.S. and Feldman, M.W. (1982) *Estuarine Comparisons* (Kennedy, V., ed.) pp. 537–551, Academic Press
34 Selander, R.K., Yang, S.Y., Lewontin, R.C and Johnson, W.E. (1970) *Evolution* 24, 402–414
35 Saunders, N.C., Kessler, L.G. and Avise, J.C. (1986) *Genetics* 112, 613–627
36 Britten, R.J., Cetta, A. and Davidson, E.H (1978) *Cell* 15, 1175–1186
37 Marcus, N.H. (1977) *Biol. Bull.* 153, 560–576
38 Koehn, R.K., Milkman, R.D. and Milton, J (1976) *Evolution* 30, 2–32
39 Levinton, J.S. and Suchanek, T.H. (1978) *Mar. Biol.* 49, 363–375
40 Rosenblatt, R.H. and Waples, R.S. (1986) *Copeia* 2, 275–284
41 Winans, G.A. (1980) *Evolution* 34, 558–574
42 Avise, J.C., Helfman, G.S., Saunders, N.C. and Hales, L.S. (1986) *Proc. Natl Acad. Sci. USA* 83, 4350–4354
43 Buroker, N.E., Hershberger, W.K. and Chew, K.K. (1979) *Mar. Biol.* 54, 157–169
44 Gooch, J.L., Smith, B.S. and Knupp, D. (1972) *Biol. Bull.* 142, 36–48
45 Hedgecock, D. (1986) *Bull. Mar. Sci.* 39, 550–564

What is a Quasispecies?

Martin A. Nowak

A quasispecies is a well-defined distribution of mutants that is generated by a mutation–selection process. Selection does not act on a single mutant but on the quasispecies as a whole. Experimental systems have been designed to study quasispecies evolution under laboratory conditions. More recently, virus populations have been called quasispecies to indicate their extensive genetic heterogeneity. The most prominent examples are probably the human immunodeficiency viruses HIV-1 *and* HIV-2. *The quasispecies nature of* HIV *has formed the basis of a model that provides a mechanism for the pathogenesis of acquired immunodeficiency syndrome* (AIDS) *in humans. This article focuses on the nature of the quasispecies concept and its implications for evolutionary biology and virology.*

The term 'quasispecies' was introduced by Eigen and Schuster[1] in 1977, in the context of their work on the origin of life, to describe the cluster of closely related molecular 'species' produced by errors in the self replication of macromolecules (nucleic acids). This followed Eigen's first theoretical model of molecular evolution based on chemical kinetics[2].

Martin Nowak is at the Dept of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

In the original notion of Eigen and Schuster, a quasispecies is defined as the equilibrium mutant distribution that is generated by a specific mutation–selection process describing the erroneous replication of macromolecules (nucleic acids)[1–7]. Suppose there are n different nucleic acid sequences I_1, I_2, ..., I_n that can serve as templates for replication. Each variant is characterized by a specific nucleotide sequence. This nucleotide sequence may determine the replication rate of a given variant. The replication rates of the variants I_1, I_2, ..., I_n may be denoted by a_1, a_2, ..., a_n. These quantities represent the selective values of the individual mutants. In the absence of mutation, the variant with the highest replication rate will grow fastest and reach fixation.

The result of selection in this world without errors is a homogeneous population consisting of the fastest replicating variant. But replication is not error free. Thus it is necessary to define the probabilities Q_{ij} that (erroneous) replication of template I_j results in the production of the sequence I_i. The quantities Q_{ij} for $i=1, 2, ..., n$ and $j=1, 2, ..., n$ form the so-called mutation matrix.

A system of ordinary differential equations describes the time evolution of the population of these nucleic acid sequences. The growth rate of a specific variant, e.g. I_1, can be written as

$$dx_1/dt = a_1Q_{11}x_1 + a_2Q_{12}x_2 + ... + a_nQ_{1n}x_n \quad (1)$$

Here x_1, x_2, ..., x_n denote the population sizes of the variants I_1, I_2, ..., I_n.

New particles of variant I_1 can be formed by error-free replication of I_1; this happens at the replication rate a_1 and the probability Q_{11}; the overall rate is therefore given by $a_1Q_{11}x_1$. Erroneous replication of any other mutants I_2, ..., I_n can also lead to new I_1 particles. This is represented by the growth terms $a_2Q_{12}x_2 + ... + a_nQ_{1n}x_n$ in Eqn 1. In the same way, we can write the rate of production of any of the other variants to obtain the whole system of differential equations

$$dx_i/dt = \sum_{j=1}^{n} a_jQ_{ij}x_j \quad i = 1, .., n \quad (2)$$

In this context, the population will no longer consist only of the fastest growing sequence, but of a whole ensemble of mutants with different replication rates. This ensemble of mutants is the quasispecies (see Box 1).

The frequency of a given variant within the quasispecies does not depend on its replicative value alone, but also on the likelihood with which it is produced by erroneous replication of other templates and their frequencies in the quasispecies distribution. This is important to the understanding of the structural organization of a quasispecies. The consequence of this effect is that the individual sequence I_i with its replicative value a_i no longer serves as the unit (or target) of selection. The quasispecies itself is the target of selection in a mutation–selection process. This fact has important implications. Evolution is normally thought of as the interaction between mutation and selection. Selection is a factor that favors advantageous mutants that have been generated by pure chance; indeed, it is normally considered a mistake to think of mutations as being guided other than by chance. A quasispecies, however, can guide mutations. This does not mean that there is any correlation between the (intrinsically stochastic) act of mutation and the selective advantage of the mutant. But selection operates on the structure of the whole quasispecies, which is adapted to its *fitness landscape* (this term is originally from Sewall Wright). Therefore, evolution can be guided towards the peaks of this fitness landscape. This happens because more-successful mutants (which may be closer to the peaks of the landscape) will produce more offspring than less-successful mutants (which may be further away from the peaks). Evolutionary optimization can be viewed as a hill-climbing process of the quasispecies that occurs along certain pathways in sequence space (see Box 2).

Error thresholds

Another important concept in quasispecies theory is the error threshold of replication[1,4–6,9]. If replication were error free, no mutants would arise and evolution would stop. Evolution would, however, also be impossible if the error rate of replication were too high (only some mutations may lead to an improvement in adaptation, but most will lead to deterioration). The quasispecies concept allows us to quantify the resulting minimal replication accuracy that maintains adaptation.

Let us assume that a population consists of (1) a fast replicating variant I_1 – the wild-type sequence – with replication rate a_1 and (2) its mutant distribution (error tail) I_2 with a lower average replication rate a_2 (see Refs 6 and 10). Let q denote the per-base accuracy of replication, i.e. the probability that a single base is accurately replicated. Thus the probability that the whole sequence (of length m) is replicated without errors is given by $Q = q^m$. Neglecting the small probability that erroneous replication of a mutant gives rise to a wild-type sequence leads to the equations

$$dx_1/dt = a_1Qx_1$$
$$dx_2/dt = a_1(1-Q)x_1 + a_2x_2$$

Here the ratio of wild type to mutants converges to

$$x_1/x_2 \rightarrow \frac{a_1Q - a_2}{a_1(1-Q)}$$

Therefore the wild type can only be maintained in the population if $Q > a_2/a_1$. This means that the single-digit replication accuracy, q, must be larger than a certain critical value. This error threshold relation is obtained as

$$q > q_{crit} = (a_2/a_1)^{1/m}$$

For replication accuracies lower than q_{crit} the wild-type sequence will be lost from the population although it has the highest replication rate. This leads to an important relationship between the replication accuracy and the sequence length

$$m < 1/(1-q)$$

Here I have used the approximation that the logarithm of a_1/a_2 is about 1. This represents an approximation

Box 1. Chemical kinetics: the origin of the quasispecies

DNA or RNA replication can be visualized as the following chemical reaction:

$$\text{error free: } (A) + I_i \xrightarrow{a_iQ_{ii}} 2I_i$$

$$\text{mutation: } (A) + I_i \xrightarrow{a_iQ_{ij}} I_i + I_j$$

The symbol A denotes low-molecular-weight materials (the four nucleotides) that are required for DNA or RNA synthesis. It is assumed that the available amount of A is constant and hence it will not enter as a variable into the kinetic differential equations. Error-free replication and mutation are parallel reactions of the same mechanism. The rate of replication, a_i, depends on the template I_i; the mutation probability, Q_{ij}, depends on both the template and the product of replication. In addition, we consider an unspecific degradation or dilution flow $I_i \rightarrow 0$, which may be adjusted in such a way that the total population is of constant size. This leads to the following differential equation describing the chemical (mass action) kinetics

$$d\vec{x}/dt = W\vec{x} - f(\vec{x})\vec{x}$$

The vector $\vec{x}$ contains the population densities of the individual sequences

$$\vec{x} = (x_1, x_2, ..., x_n)$$

The matrix W contains the replication rates and mutation probabilities

$$W = \begin{pmatrix} a_1Q_{11} & a_2Q_{12} & \cdots & a_nQ_{1n} \\ a_1Q_{21} & a_2Q_{22} & \cdots & a_nQ_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_1Q_{n1} & a_2Q_{n2} & \cdots & a_nQ_{nn} \end{pmatrix}$$

The total population size, Σx_i, remains constant if

$$f(\vec{x}) = \sum_{i=1}^{n} a_ix_i \Big/ \sum_{i=1}^{n} x_i$$

The equilibrium of Eqn 2 can be calculated by solving the standard eigenvalue problem of linear algebra

$$W\vec{x} = \lambda\vec{x}$$

Now the quasispecies can be defined – in precise mathematical terms – as the dominant eigenvector $\vec{x} = (x_1, x_2, ..., x_n)$ which belongs to the largest eigenvalue λ_{max} of the matrix W. This eigenvector, $\vec{x}$, describes the exact population structure of the quasispecies; each mutant I_i is contained in the quasispecies with frequency x_i. (We can normalize such that $\Sigma x_i = 1$.) The largest eigenvalue is exactly the average replication rate of the quasispecies, $\lambda_{max} = \Sigma a_ix_i$.

Box 2. The sequence space: short distances, many routes

Nucleic acids such as DNA or RNA consist of specific sequences (or strings) of the four individual nucleotides (G, A, C and U or T). For a given chain length m there are 4^m different variants. This means that even for moderate chain lengths a 'hyper-astronomically' large number of different variants can be formed. For example, for a polynucleotide of length $m=300$, which is just large enough to encode one of the smallest proteins, there are more than 10^{180} different variants. The genome length of HIV is about 10 000 bases. A particular HIV sequence is one choice out of 10^{6020} different nucleotide sequences of the same length. For comparison, there are only 10^{11} stars in our galaxy, or about 10^{80} protons in our universe; Avogadro's number is only 6×10^{23}.

Let us imagine all possible nucleotide sequences to be arranged in a 'sequence space' such that two sequences are neighbors if one can be converted into another by a single point mutation[8]. Thus the sequence space is formed by a set of sequences (of uniform length m) together with a definition of a distance between sequences. An appropriate definition is given by the Hamming distance. Let us specify two sequences I_i and I_j; the Hamming distance counts the number of different positions in these two sequences. It represents the total number of point mutations that are necessary to change one sequence into the other. A correct ordering of all the mutants according to their mutual Hamming distances leads to an m-dimensional space. The important features of this sequence space are (1) its high dimensionality, (2) the large number of shortest mutational routes between two distant mutant sequences (for two sequences separated by a Hamming distance d there are $d!$ shortest mutational routes) and (3) that many sequences are confined to a close neighborhood of each other. The diameter of a sequence space that contains 10^{80} points is only 133 length units, i.e. point mutations. This means that relatively few point mutations can lead from one region in the sequence space to a completely different region, providing there exists something like a guiding gradient to avoid going in 'wrong directions'. In evolution this gradient is provided by natural selection.

for the upper genome length m that can be maintained by a given single-digit replication accuracy without losing adaptation.

Experimental studies

The quasispecies concept becomes important whenever mutation rates are high. This is often the case in viral and bacterial populations. The first '*in vitro* darwinian evolution experiment' was performed by Sol Spiegelman and colleagues[11]. In this traditional serial transfer experiment, adaptation of the *E. coli* phage Qβ to an artificial selection pressure was demonstrated. The artificial selection pressure was exerted by ethidium bromide, which inhibits (slows down) replication. After a certain number of passages, a mutant sequence evolved that was growing twice as fast as the original wild type in the presence of ethidium bromide. The difference between wild type and mutant was just three point mutations. More recently, machines have been constructed to perform these serial transfers automatically. Such 'evolutionary reactors' allow one to study adaptation of viruses to a given (artificial) selection pressure (H. Otten, Thesis, Max Planck Institut, Göttingen, 1990).

Biebricher *et al.*[12] studied the quasispecies distribution of short-chained RNA templates in a cell-free medium with Qβ replicase. Sequence analysis indicated a broad distribution of mutants around a master sequence. In this system, mutation rates and selective values of individual sequences can be measured.

Another interesting observation was the *de novo* synthesis of RNA templates by the Qβ replicase[13,14]. Initially the system contains only Qβ replicase and nucleotide triphosphates, but no RNA molecules. Some templates are then formed by chance. These short templates compete for recognition and replication by the Qβ enzyme. This leads to the outgrowth of the best-adapted sequence.

Mutation rates can be determined experimentally. Mutation frequencies (i.e. the probability that a replication enzyme makes a misincorporation at a certain position) have to be distinguished from mutant frequencies (i.e. the proportion of certain mutants in a population). For example, a 'hot spot' is a region in the genome with very high mutant frequency. A hot spot can be generated in two ways: (1) the mutation rate at this position is very high or (2) selection favors (or tolerates) variation in this region. In the second case, mutant and mutation frequencies are completely different. Thus, to count mutants in a quasispecies distribution is not a way to determine mutation rates. An elegant method to obtain mutation frequencies is to measure the reversion rate of conditionally lethal mutants (produced by site-directed mutagenesis).

Viral quasispecies

Virus populations in gener consist of a widely disperse mutant distribution rather than homogeneous population of single wild-type sequence[15–17]. Erro rates have been determined, fo example, for influenza A virus[1 vesicular stomatitis virus[19], foot and-mouth disease virus[20], splee necrosis virus[21,22] and HIV-1[23,24]. Al results show a correlation betwee error rate and sequence length.

In a classical experiment by Domingo *et al.*[25] single particles o the phage Qβ were cloned. None o the different clones agreed exactly in their genomic sequences. The actual 'wild type' was present at a level of less than 5%.

The human immunodeficiency virus is an important example of a viral quasispecies. Patients infected with HIV harbor a highly diverse virus population with many different mutants[26–29]. Mutations are generated by the virus-encoded reverse transcriptase, which has an error rate of about 10^{-4} to 10^{-3} per base. This implies that during each replication of the whole genome 1–10 errors are produced. HIV seems to operate very close to its error threshold[30]. These observations have led to the formulation of a mathematical theory that outlines the potential importance of antigenic variation as a major factor driving disease progression[31–33]. The essential idea is that the virus evades immune pressure by the continuous production of new mutants resistant to current immunological attack. This results in the accumulation of antigenic diversity during the asymptomatic period. The existence of an antigenic diversity threshold is derived from the asymmetric interaction between the virus quasispecies and the immune system: strain-specific immune responses are directed against specific HIV antigens, but each virus strain can impair all immune responses regardless of their specificity. Therefore, increasing HIV diversity enables the virus population to escape from control by the immune system.

In this context, the observed genetic variability is responsible for the fact that the virus establishes a persistent infection without being cleared by the immune response

and induces immunodeficiency disease after a long and variable incubation period.

HIV can also evolve drug resistance after about six months of treatment. Resistance against zidovudine, the drug most widely used to treat HIV infections, is mediated by three or four amino acid substitutions in the virus-encoded reverse transcriptase[34].

Conclusions

Quasispecies theory, which is based on chemical kinetics, provides a mathematical framework in order to understand molecular evolution. Selection and mutation form a distribution of mutants that is called a quasispecies. The target of selection is not an individual mutant sequence but the whole quasispecies. Therefore, fitness is a property of the quasispecies and not of individual mutants.

The fitness of a quasispecies is mathematically defined as the largest eigenvalue of the mutation–selection matrix. Selection stabilizes a quasispecies distribution in sequence space, and evolution can be viewed as a destabilization of an existing quasispecies upon arrival of a new advantageous mutant that establishes a new quasispecies[35,36].

A quasispecies may be centered around a master sequence with high efficiency of reproduction. The consensus sequence of the quasispecies need not be identical with the master sequence. For larger mutation rates, the frequency of the master sequence in the population can be very low. Evolution seems to work fastest close to the error threshold. The quasispecies has changed the classical view of evolution from the picture of a single wild type moving through sequence space by random walk[8] into the picture of a quasispecies with its mutant distribution migrating through sequence space in an internally self-controlled manner and guiding itself to the peaks of the fitness landscape.

References

1 Eigen, M. and Schuster, P. (1977) *Naturwissenschaften* 64, 541–565
2 Eigen, M. (1971) *Naturwissenschaften* 58, 465–526
3 Fontana, W. and Schuster, P. (1987) *Biophys. Chem.* 26, 123–147
4 Swetina, J. and Schuster, P. (1982) *Biophys. Chem.* 16, 329–345
5 McCaskill, J.S. (1984) *J. Chem. Phys.* 80, 5194–5204
6 Nowak, M. and Schuster, P. (1989) *J. Theor. Biol.* 137, 375–395
7 Eigen, M., McCaskill, J.S. and Schuster, P. (1989) *Adv. Chem. Phys.* 75, 149–263
8 Maynard Smith, J. (1970) *Nature* 225, 563–564
9 Szathmáry, E. (1989) *Trends Ecol. Evol.* 4, 200–204
10 Maynard Smith, J. (1989) *Evolutionary Genetics*, Oxford University Press
11 Mills, D.R., Kramer, F.R. and Spiegelman, S. (1973) *Science* 180, 916–918
12 Biebricher, C.K., Eigen, M. and Gardiner, W.C. (1985) *Biochemistry* 24, 6550–6574
13 Semper, M. and Luce, R. (1975) *Proc. Natl Acad. Sci. USA* 72, 162–166
14 Biebricher, C.K., Eigen, M. and Luce, R. (1986) *Nature* 321, 89–92
15 Domingo, E. *et al.* (1986) *Gene* 40, 1–37
16 Holland, J. *et al.* (1982) *Science* 215, 1577–1582
17 Wain-Hobson, S. (1989) *AIDS* 3, s13–s18
18 Parvin, J.D., Moscona, A., Pan, W.J., Lieder, J. and Palese, P. (1986) *J. Virol.* 59, 377–398
19 Spindler, K.R., Horodyski, F.M. and Holland, J.J. (1982) *Virology* 119, 96–128
20 Domingo, E., Darita, M. and Ortin, J. (1980) *Gene* 11, 333–357
21 Pathak, V.K. and Temin, H.M. (1990) *Proc. Natl Acad. Sci. USA* 87, 6019–6023
22 Pathak, V.K. and Temin, H.M. (1990) *Proc. Natl Acad. Sci. USA* 87, 6024–6028
23 Preston, B.D., Poiesz, B.J. and Loeb, L.A. (1988) *Science* 242, 1168–1171
24 Roberts, J.D., Bebenek, K. and Kunkel, T.A. (1988) *Science* 242, 1171–1173
25 Domingo, E., Sabo, D., Taniguchi, T. and Weissmann, C. (1978) *Cell* 13, 735–762
26 Saag, M.S. *et al.* (1988) *Nature* 334, 440–444
27 Meyerhans, A. *et al.* (1989) *Cell* 58, 901–910
28 Simmonds, P. *et al.* (1990) *J. Virol.* 64, 5840–5853
29 Hahn, B.H. *et al.* (1986) *Science* 232, 1548–1553
30 Nowak, M.A. (1990) *Nature* 347, 522
31 Nowak, M.A., May, R.M. and Anderson, R.M. (1990) *AIDS* 4, 1095–1103
32 Nowak, M.A. *et al.* (1991) *Science* 254, 963–969
33 Nowak, M.A. *J. Theor. Biol.* (in press)
34 Larder, B.A. and Kemp, S.D. (1989) *Science* 246, 1155–1157
35 McCaskill, J.S. (1984) *Biol. Cybern.* 50, 63–75
36 Fontana, W., Schnabl, W. and Schuster, P. (1989) *Phys. Rev. A* 40, 3301–3321

Letters to the Editor

TREE welcomes correspondence. Letters to the Editor may address issues raised in the pages of *TREE*, or other matters of general interest to ecologists and evolutionary biologists. Letters should be no more than 300 words long, and should be sent to The Editor, *Trends in Ecology and Evolution*, Elsevier Trends Journals, 68 Hills Road, Cambridge CB2 1LA, UK. The decision to publish rests with the Editor, and the author(s) of any *TREE* article criticized in a Letter will normally be invited to reply.