From: Evolution in Health and Disease, edited by Stephen C. Stearns, Oxford University Press, 1999

2

RESEARCH DESIGNS THAT ADDRESS EVOLUTIONARY QUESTIONS ABOUT MEDICAL DISORDERS

Randolph M. Nesse and George C. Williams

Diseases result usually from webs of interacting causes of enormous complexity, while human minds seek explanatory principles of extraordinary simplicity. This conflict gives rise to a central problem for medicine. Explanations of disease, and most programmes of medical research, tend to emphasize a single cause, while most diseases result from multiple environmental factors interacting with several sources of vulnerability. An evolutionary approach fosters clear thinking about the complex origins of disease. It is, however, easy to underestimate the change in perspective that an evolutionary view of disease offers and requires. Many evolutionary research questions differ qualitatively from those usually pursued in medicine, and the routine application of traditional modes of medical inference may prove inadequate to the task of testing them. Addressing the formulation and testing of evolutionary hypotheses may seem like abstract philosophy, but it is essential if this enterprise is to succeed.

New evolutionary perspectives on basic mechanisms are leading to substantial discoveries in established areas of medical research, such as virulence, senescence, and genetic variation, as documented by the chapters in this book. Equally important is the potential for integrating multiple causes of specific diseases provided by an evolutionary approach. The benefits of this approach will be delayed, however, if its initial applications are poorly done. It is easy to make up evolutionary stories about adaptation but harder to test them. This chapter outlines how evolutionary hypotheses about specific diseases can be formulated and tested. It is premature to promote any such system as definitive, and we hope that readers will sympathize with the difficulties of trying to impose some order, however preliminary and arbitrary, on an unruly tangle of questions.

This chapter proceeds in two steps. First, we list several distinctions that are essential to defining the objects of explanation and the kinds of explanations proposed, and second, an outline of how hypotheses about vulnerability to diseases can be formulated and the kinds of evidence that can be used to test them. Examples are provided in other chapters; Chapter 23 applies this framework systematically to the major psychiatric disorders.

DISTINCTIONS FOR DARWINIAN DOCTORS

The most common and serious difficulties in pursuing an evolutionary perspective on disease arise from lack of clarity about fundamental distinctions (Table 2.1). While some of them will be familiar, this outline will be worthwhile if it spares some unnecessary misunderstanding and argument.

The distinction between proximate and evolutionary explanations for an adaptive trait is critical. A proximate explanation describes a trait's ontogeny and the resulting anatomy and physiology and how they work to accomplish the trait's function. An evolutionary explanation, by contrast, describes a trait's phylogeny and how the trait gives a fitness advantage

Table 2.1 Some distinctions for Darwinian doctors

1. Proximate versus evolutionary explanations
2. Explanations of individual differences (why some individuals get a disease when others do not) versus traits individuals share (why all members of a species are vulnerable to a disease)
3. Manifestations of disease that arise from defects versus defences (or other adaptations)
4. Explanations of a disease itself versus explanations of why the body is designed so it is vulnerable to a disease
5. Group versus individual versus gene levels of selection
6. Most traits are involved in tradeoffs that force compromises
7. The different meanings of adaptation, fitness, defence, and evolution in different fields
8. The several distinct kinds of evolutionary explanations for vulnerability to disease

(Mayr 1982). Tinbergen (1963) clarified the distinction when he listed four questions that must be answered to explain any adaptive trait completely:

1. What is the proximate mechanism of the trait-its structure and operation-at all levels from chemistry, to anatomy, to physiology and interactions with other traits?
2. What is the ontogeny of the trait, from the zygote to the mature individual?
3. How does the trait give a fitness advantage that can account for selection for the genes that give the trait its current form? What adaptive functions does it serve and how do those functions increase Darwinian fitness?
4. What is the phylogeny of the trait? What are its precursors and what forces and intermediate stages gave rise to its present form?

The answers to the first two questions are both parts of the proximate explanation; the answers to the last two questions are both parts of an evolutionary explanation. Proximate and evolutionary explanations are distinct and complementary; both are necessary for a full understanding of any trait shaped by natural selection.

Proximate explanations seem more straightforward than evolutionary explanations because they refer to observable physical structures and because often only one hypothesis can be correct. DNA either is or is not a double helix. Cortisol either does or does not come from the adrenal cortex. Evolutionary explanations, by contrast, refer to past events whose traces are found in fossils, in gene frequencies, and in the resulting body structures, and they refer to mechanisms of natural selection involving competition of phenotypes and traits within populations.

The obvious difficulties of investigating past events and population processes are compounded by the possibility of multiple explanations. Eyebrows may both keep sweat out of the eyes and signal conspecifics. The tongue may both process food and form words. Beyond these difficulties, still other problems arise. What qualifies as a trait? How can we tell if we are carving the body at nature's joints? How can we tell if a trait is an adaptation shaped by natural selection or something else? The temptation is to give up, but the problems are important. Solving them is not impossible, just difficult. Physiology has a well-developed tool kit for determining the functions of an organ: observing the effects of extirpation, determining if form matches function, natural experiments, and more (Vander et al. 1985; Schmidt-Nielsen 1990). Behavioural ecology also has well-developed methods (Krebs and Davies 1997). A comparable tool kit for evolutionary medicine is still being assembled.

A second crucial distinction is between explanations of why one individual gets sick and another does not, versus an explanation of why all members of a species share some vulnerability to disease. The reasons why some individuals on early ship voyages developed scurvy long before others are different from the reasons why all humans require vitamin C. Explaining why some members of modern populations will develop disease from arterial occlusion and others will not is a problem separate from that of explaining why atherosclerosis is so common. Medical research has focused, understandably, on trying to explain

why one person beomes sick and another does not. Such questions differ fundamentally from evolutionary questions about why all humans are vulnerable to certain diseases.

A third distinction is between manifestations of disease that reflect defects and those that are aspects of defences. Many medical problems, like seizures, jaundice, and hallucinations arise from defects in the body's machinery. Others, like fever, cough, pain, and anxiety, are defensive responses shaped by natural selection that remain latent until aroused by a threat. Because they are usually unpleasant and associated with an unfavorable situation, it has been easy to confuse these defences with the diseases that arouse them. But as we develop new drugs that block defences more effectively, we need to recognize the functions of these defences and understand when it is, and is not, safe to block them.

A fourth distinction is between explanations of a disease and aspects of the body that make us vulnerable to it. Diseases are, in general, not shaped by natural selection and have no adaptive explanation. But diseases arise from some vulnerability in the body's design. These design features are shaped by natural selection, directly or indirectly, and they do have evolutionary explanations. This small change in focus, from a disease itself, to the aspects of the body that make it vulnerable to disease, makes the evolutionary approach to medicine viable.

A fifth distinction is between levels at which selection may operate, especially between the gene, the individual, and the group. Despite vivid expositions on selfish genes (Dawkins 1976), many physicians remain unaware of the difficulty with explanations based on naive group selection (Williams 1966). The main difficulty is that selfish mutants, or selfish phenotypes, continually arise and invade populations of organisms that are sacrificing their reproductive interests for the good of some larger unit. It takes rather special and implausible circumstances for a refined group-selection theory (e.g. Wilson and Sober 1994) to work at all. Among those who grasp the possibility that genes can persist at the expense of individuals, many imagine that mutations in somatic cells can somehow help themselves by inducing cancerous cell division. In fact, of course, an individual's somatic cells are, for the most part, genetically identical and so they do not have conflicting interests (with a few possible exceptions), while different individuals have different genes, and therefore different interests that give rise to conflicts.

Related individuals do, of course, have some proportion of their genes in common, so if a gene decreases an individual's reproductive success but increases that of relatives (say, for instance, a gene that tends to make parents protect their children) then the frequency of the gene can increase by kin selection-so long, that is, as the cost to the actor is less than the benefit to the relatives multiplied by the percentage of genes identical by descent (Hamilton's rule, of Hamilton 1964). Many people, on first grasping kin selection, are initially distracted by Washburn's fallacy, thinking that common interests are based on the percentage of genetic code that is identical, say the 97 per cent of the nucleotide sequences that humans and chimpanzees have in common. In fact, the relevant factor is the percentage of genes in common by immediate genealogy, 50 per cent for first-degree relatives, 25 per cent for second-degree relatives, and so forth (Williams 1996: 43-51).

Sixth, selection can influence gene frequency in several ways. The traditional dichotomy between natural selection and sexual selection conceals their underlying common mechanism, and the different ways that effects of genes can increase their representation in future generations (Andersson 1994). The trade-offs between these different 'strategies' are at the core of understanding how natural selection shapes the organism (Stearns 1992; Krebs and Davies 1997). Survival is essential, but it is selected only in the service of direct and indirect reproduction. When a gene increases net reproductive success at the expense of individual health or survival, whether by effects on sexual competition, display, fertility, or parenting, that gene will tend to increase in frequency. This may underlie the shorter lifespan

of men than of women. Also important are trade-offs between mating effort and parenting effort, between investing in self and in kin, and between investing much in a few offspring versus less in more offspring.

Seventh, one must carefully specify the meanings of several words that are used differently in different fields. Darwinian fitness, or reproductive success through oneself and one's relatives, is quite different from everyday meanings of fitness and health. Adaptation in biology is very different from the several meanings of adaptation in psychology or neurology. An inducible defence, in biology, is a special protective trait or a state that is evoked by certain threats, a very different matter from the notion of defence in psychoanalysis, and only somewhat similar to the idea of defence in war, sports, or law. Even the word evolution itself most often refers in medical databases not to the process of biological descent with modification, but to all manner of gradual changes in organisms, organizations, and beliefs.

Finally, it is essential to distinguish the main possible kinds of evolutionary explanations for vulnerability to disease (Williams and Nesse 1991; Nesse and Williams 1994: 8-11). These categories (Table 2.2 and Fig. 2.1) give us a foundation for analysing vulnerability to disease.

Table 2.2 Categories of evolutionary explanation for vulnerability to disease

1. Defence-what we think is a disease or defect is actually an adaptation
2. Conflict with other coevolving agents, such as pathogens
3. Novel aspects of the current environment
4. Genetic quirks that are harmful only in a novel environment
5. Design trade-offs at the level of the gene
6. Design trade-offs at the level of the trait
7. Historical legacy and path dependence
8. Random factors

POSSIBLE EXPLANATIONS FOR VULNERABILITY TO DISEASE

Each of the categories in Table 2.2 generates hypotheses that are somewhat different. These hypotheses, in turn, require different tests that pose different special problems. Much of the confusion in evolutionary approaches to disease arises, we believe, from lack of clarity about the exact kind of evolutionary hypothesis being considered and the different ways in which different kinds of hypotheses can be tested. Each category is addressed in turn, with a brief description, followed by the most characteristic predictions (formulated as 'If . . . then . ..' statements), and a note, if necessary, about special problems that arise when testing hypotheses in each category.

The first possibility is that an adaptation has been misinterpreted as a defect, usually because it is a defence against some threat. Inducible defences are readily confused with defects and with causes of diseases, for they are often unpleasant and are manifested mainly in association with disease. Examples include fever, cough, pain, vomiting, anxiety, and jealousy. Dysregulation of defences often results in disease; examples include autoimmune diseases, pulmonary embolism, seizures from high fever, anxiety disorders, and depression.

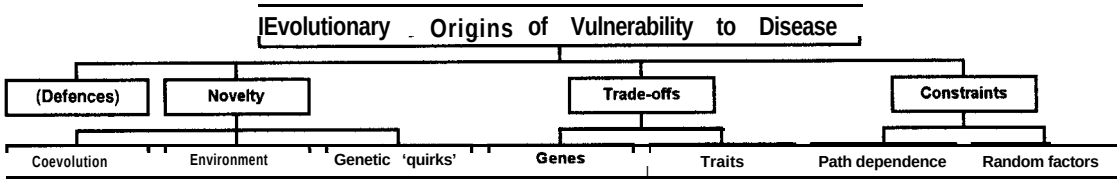

Fig. 2.1 Evolutionary origins of vulnerability to disease.

Hypotheses about defences assert that a trait is an evolved adaptation of a special kind; testing them involves all the complexities and judgement required to assess the status of a trait as an adaptation (Rose and Lauder 1996: Part 1). No uniform standards of evidence exist for making such assessments, but when hypotheses are formulated as 'if trait T is a defence, then . ..'. several kinds of predictions can be made (Table 2.3).

Testing hypotheses about defences involves still other inherent difficulties. First, looking for the utility of defences is counterintuitive because they usually are unpleasant, seem harmful, and are observed only in conjunction with danger. Second, they usually are somewhat harmful, otherwise they would be expressed all the time. Third, because many defences are inexpensive compared with the harms they protect against, selection shapes regulation mechanisms, according to the principles of signal detection theory, that tolerate many false alarms to ensure a response when it is essential (the 'smoke-detector principle', see Nesse and Williams 1994: 159). Because of this, we often see no ill effects when drugs block defences, and this fosters the illusion that the defences are not useful. Finally, in the relative safety of the modern environment, many defences are excessive. For some, like the fight-flight response, most instances of expression are now pathological (panic attacks, destructive violence). All of this has very practical implications as new drugs that block aversive emotions are discovered and widely used.

Table 2.3 If trait T is a defence, then:

1. Individual differences in the degree of expression of the trait should influence the degree of protection against the threat
   (a) Inherited defects
   (b) Inherited variation
   (c) Surgical excision
   (d) Drug manipulation
2. The form of the trait should match its function- details of the state should match those that defend against a threat
3. If there are subtypes of the defence, their characteristics should match the demands posed by specific threats, and the regulation system should express the correct subtype in the face of different threats
4. Details of the trait's proximate mechanisms should match expectations based on its function
5. Phylogenetic relatives should have homologous adaptations that vary according to the kinds and amounts of the threats they experience
6. Unrelated species that face similar threats may have analogous adaptations
7. The mechanisms that regulate the expression of the defence should express the kind, degree, and duration of response that is appropriate to the threat
8. Facultative mechanisms may increase the sensitivity of the regulation system in individuals who are more exposed to the threat
9. The proximate mechanisms of the regulation system should match the challenges posed by the threat

A recent example of a test of the idea that an evolved feature is a defence is the proposal that menstruation serves to rid the reproductive tract of infections introduced by sperm (Profet 1993). Profet's arguments were based largely on category 4 tests in Table 2.3, but she also made or implied other predictions, for instance in category 5. She noted that danger from pathogens introduced by semen should be greatest in those species with the most frequent mating and greatest diversity of mates. That the most profuse menstruation is found in women and female chimpanzees accords nicely with this expectation. Unfortunately for Profet's theory, a more inclusive phylogenetic analysis of menstrual phenomena in relation to mammalian mating systems, recently carried out by Strassmann (1996) showed that Profet's expectations are not borne out.

Strassmann's work clearly shows that defence against infections of the reproductive tract is not the reason why menstruation exists, but this does not mean that it has no value as a defence against infection. Evolution frequently modifies extant features to serve new functions. The reason birds have wings is not the same as the reason vertebrates have fore-

limbs. Some of the features that Profet noted, for instance differences between menstrual and circulating blood, might well relate to defence against pathogens, and if so, they are of medical importance. Strassmann's alternative to Profet's theory is that menstruation is a special mammalian aspect of a general vertebrate adaptation for energy conservation. Keeping the reproductive tract constantly ready for use, even when no use is possible, is more expensive than intermittently shutting it down, even when the shutting down requires considerable loss of blood. We have recently learned of impending challenges to both Profet's and Strassmann's ideas. The adaptive significance of menstruation is a topic for which such attention is long overdue.

A second possible explanation is conflict with other organisms. Pathogens evolve faster than their hosts, ourselves, and have evolutionary interests that are often in conflict with ours. Thus pathogens pose a constantly novel environment in which our adaptations are always out of date (Ewald 1994). Parent-offspring conflict (Trivers 1974) is a related example of importance to psychiatry (Slavin and Kriegman 1992) and a special type of parent-offspring conflict, maternal-fetal conflict, is of potential major medical importance (Haig 1993 and Chapter 7). Explanations invoking conflicts are covered in detail in other chapters. Here we note only that such arms races can cause disease in apparently unrelated systems; for instance, heart damage caused by rheumatic fever because of immune responses to a streptococcus that mimics human antigens. Also, the social environment may involve arms races as challenging as those that result from competition with microbes, and may create designs of remarkable complexity and fragility, perhaps shaping human tendencies for mental conflict and psychopathology (Alexander 1987; McGuire *et al.* 1992).

A third explanation is the *mismatch* between our bodies and novel aspects of the environment; novel, that is, since the development of agriculture, and especially since the industrial revolution. Natural selection cannot change the body's design fast enough to cope with rapid environmental changes. The resulting diseases include atherosclerosis, breast cancer, substance abuse, eating disorders and probably depression. If a disease is caused by a mismatch between our bodies and novel factors in the environment, then:

(1) groups more exposed to the novel factor should have higher rates of the disorder;

(2) the disease should be much more prevalent in modern than ancestral environments;

(3) within groups, individuals more exposed to the factor should have increased rates of the disorder; and

(4) it should be possible to demonstrate the proximate mechanism by which the factor causes disease.

Fourth, diseases that result from novel environmental factors may affect only some individuals because of differences in exposure and because of genetic variation. *Genetic quirks* that vary between individuals and cause no harm in the ancestral environment may dramatically increase the risk of disease in modern environments. It makes no sense to call these variations 'defects', for they were harmless, or even helpful, until the past few thousand years. Examples include genetic factors in diabetes, coronary artery disease, depression, and substance abuse. If a disease is postulated to result from a genetic quirk, then:

(1) individuals with the responsible allele will have normal reproductive success in the ancestral environment, but decreased health in the modern environment;

(2) the degree of decreased fitness among those with the allele will be related to the degree of exposure to the novel factor; and

(3) individuals without the allele will have normal phenotypes in both ancestral and modern environments.

A fifth explanation is trade-offs at the level of the gene. Some individuals have alleles that give them fitness advantages (at least in some environments or genotypes) but at the cost of increased vulnerability to disease.

Manic depressive illness will be addressed in this perspective. Often the frequency of such genes varies between groups because different environments pose different threats. The haemoglobinopathies that protect against malaria are the best known examples. Most mutations that confer an advantage will also give disadvantages. If selection pushes such a gene to fixation, all members of a species may be vulnerable to a disease. If the disease is prevalent and severe enough, there will be further selection for modifier genes that decrease the adverse effects. The complexity of the resulting interaction is probably responsible for maintaining considerable genetic diversity and susceptibility to disease (Sing and Hanis 1993). In the simplest case, if a disease is caused by a gene that was selected for despite causing disease (such as the gene that causes cystic fibrosis or, perhaps, manic depressive illness), then:

(1) it may be possible to identify the specific benefit that confers this advantage;

(2) individuals with the gene should have a net inclusive fitness advantage in some environments (unless the allele is maintained by segregation distortion); and

(3) the mechanism by which the gene confers the advantage should be explicable.

The sixth explanation is trade-offs at the level of complex traits. The design of every trait is constrained by the effects on other traits. If a trait gives rise to disease because of a trade-off with other costs and benefits, then alternative designs that seem superior will, in fact, result in an overall decrement in fitness. This applies to every trait in every organism and helps to explain why nearly everything that can go wrong in a body sometimes does go wrong. A simple example is the limited strength of our hands, the price of dexterity. For mental disorders, Chapter 23 will consider the advantages and disadvantages of anxiety versus risk taking.

Seventh, historical constraints result from path dependence. Modern technological examples of such problems are vividly described in the phrase 'why things bite back' (Tenner 1996). The design for some fundamentally flawed traits cannot be corrected, because forms intermediate between the current design and a superior one have decreased fitness. The inside-out design of the vertebrate eye is an example. Human patterns of attachment and cognition may be others. If vulnerability to a disease is hypothesized to result from historical constraints, then forms intermediate between the current design and a superior alternative should give lower fitness.

Finally, random factors affect the gene pool and developmental processes. Gene frequencies are determined both by selection and by random factors including mutation, drift, and bottlenecks. These random factors are sufficient to explain most rare genetic disease, but the simple model of mutation balanced by selection is often applied to common genetic diseases where it is may well not apply. At what point does a gene's frequency, penetrance, and effects on fitness suggest that we should begin to look for selective forces that might maintain its frequency? The answers are very much up in the air (Chadwick and Cardew 1996); the issue is addressed by several chapters in this volume. If a disease results from genes that are present for reasons other than natural selection, then the balance between rates of mutation and the force of selection in the ancestral environment should be consistent with the persistence of the genes, except in the presence of special factors like segregation distortion.

These eight categories clarify hypotheses about the different ways natural selection can explain vulnerability to disease and how they can be tested. Addressing each kind of evolutionary explanation separately clarifies how they can combine to explain a specific disease.

SUMMARY

Evolutionary biology is providing new insights into the specific mechanisms of disease, including host-pathogen contests, genetic variation, and senescence. It also offers a framework for understanding why the body is vulnerable to specific diseases. Applying evolutionary

approaches to medical problems involves questions about adaptation for which canons of evidence are not well established. Progress will be possible, however, if we can specify precisely what is to be explained, distinguish the possible evolutionary explanations, and formulate and test hypotheses appropriate to each kind of explanation.