SCIENCE AND SOCIETY

From *The Origin of Species* to the origin of bacterial flagella

Mark J. Pallen and Nicholas J. Matzke

Abstract | In the recent Dover trial, and elsewhere, the 'Intelligent Design' movement has championed the bacterial flagellum as an irreducibly complex system that, it is claimed, could not have evolved through natural selection. Here we explore the arguments in favour of viewing bacterial flagella as evolved, rather than designed, entities. We dismiss the need for any great conceptual leaps in creating a model of flagellar evolution and speculate as to how an experimental programme focused on this topic might look.

Almost all microbiologists are familiar with either the bacterial flagellum (BOX 1) or type III secretion systems (T3SS), but most would have been surprised to hear these subjects feature prominently in a United States courtroom. Yet that is precisely what happened last year in the Kitzmiller versus Dover trial in Pennsylvania, where the term 'flagellum' and its cognates appeared 385 times in the transcripts of the 6-week trial. These topics were brought to the attention of the public, as the trial judge heard quotations from eminent bacteriologists, such as David DeRosier, Carl Woese, Richard Lenski, Richard Losick and Lucy Shapiro. The chief findings of the trial judge, John E. Jones III, can be summarized by a few sentences: he found it "…abundantly clear that the board's 'intelligent design' (ID) policy violates the establishment clause. In making this determination, we have addressed the seminal question of whether ID is science. We have concluded that it is not, and moreover that ID cannot uncouple itself from its creationist, and thus religious, antecedents." In short, the 'Bacterial Flagellum Trial' (as one of the defence lawyers called it) established that the teaching of ID in American state schools was unconstitutional.

The myth of irreducible complexity

But why was microbiology discussed in this court case? The answer lies in the fact that members of the ID movement (including two witnesses for the defence in the Dover trial, Michael Behe and Scott Minnich) cite the bacterial flagellum as an example of an irreducibly complex system. This complexity, they argue, could not have arisen through gradual variation and natural selection, which were proposed as the mechanism of evolution in Darwin's *The Origin of Species* (BOX 2). At the heart of the ID argument is the supposition that some biological systems are so complex that they can only function when all of their components are present, so that the system could not have evolved from a simpler assemblage that did not contain the full machinery — essentially a modern re-working of the old creationist argument 'what good is half a wing?'

Kenneth Miller, who appeared as a witness for the plaintiffs, elaborated, in non-technical terms, some of the arguments against the notion that the flagellum is irreducibly complex (see Further information for links to trial material); he and others have also done so in print[1,2]. Crucially, Miller pointed out that the flagellum is modular, in that the T3SS that is responsible for flagellar protein export constitutes a functionally intact subsystem capable of performing a useful function (protein secretion) in the absence of the rest of the flagellar apparatus. However, there are additional arguments, which we elaborate below, in favour of viewing bacterial flagella as evolved — rather than designed — entities.

Beyond typology

As the great evolutionist Ernst Mayr noted, one of Darwin's greatest achievements was to abolish typological or essentialist thinking from biology; instead, the emphasis in biology is on variation and individuality[3]. Therefore, when discussing flagellar evolution it is important to appreciate that there is no such thing as 'the' bacterial flagellum. Instead, there are myriad different bacterial flagella, showing extensive variation in form and function. The most well-studied bacterial flagellar system is that of *Salmonella enterica* serovar Typhimurium (*Salmonella typhimurium*) (BOX 1). However, in Gram-positive bacteria, flagella lack P- and L-rings[4], and in spirochaetes the flagellar filaments rotate inside the periplasm[5]. Some flagella rotate using proton-motive force, others depend on a sodium-ion gradient[6,7]. In *Sinorhizobium meliloti* and *Rhodobacter sphaeroides*, the flagellum rotates unidirectionally, with a fast, slow or stop mechanism, whereas in *S. typhimurium* reversals in the direction of flagellar rotation are used to re-orientate the cell[8,9]. Flagellar filaments vary in their physical properties: some show right-handed helical packing, others left-handed; some are flexible, others rigid; some are straight, others curly[10]; some undergo post-translational modifications such as glycosylation or methylation, others do not[11,12]. In *Escherichia coli* alone there are over forty antigenically distinct flagellins, with good evidence that variation is driven by diversifying natural selection[13]. Indeed, flagellins in general provide a perfect illustration of Darwin's dictum[14] that "nature is prodigal in variety, though niggard in innovation", in that the surface-exposed domains are highly variable, ranging in length from effectively zero to 800 residues, yet the peripheral domains that mediate inter-subunit interactions are highly conserved[15]. Furthermore, some systems deploy a single flagellin, whereas others, like the *S. typhimurium* model system, exploit two different flagellins, but never in the same filament. In other exceptional cases, up to six different flagellins are incorporated into a single flagellum[16].

Many new flagellar systems have been discovered through genome sequencing — a trend that is likely to increase with time. For example, over three hundred flagellin sequences were obtained in a single sequencing project that focused on samples from the Sargasso Sea[17]. By even the most conservative estimate, there must therefore be thousands of different bacterial flagellar systems, perhaps even millions. Therefore, there is no point discussing the creation or ID of 'the' bacterial flagellum. Instead, one is faced with two options: either there were thousands or even millions of individual creation events, which strains Occam's razor to breaking point, or one has to accept that all the highly diverse contemporary flagellar systems have evolved from a common ancestor.

Another line of evidence suggesting that flagellar systems are subject to the same evolutionary forces as other biological entities comes from the discovery of vestigial

systems. Darwin himself cited vestigial organs as evidence for evolution[14] (or as he called it, descent with modification): "On the view of descent with modification, we may conclude that the existence of organs in a rudimentary, imperfect and useless condition, or quite aborted, far from presenting a strange difficulty...might even have been anticipated and can be accounted for by the laws of inheritance." Natural selection has clearly rendered flagellar systems non-functional when they are no longer needed — vestigial non-functional remnants of flagellar genes or regulons have been discovered in several bacterial species[18–21]. This phenomenon is evident even in the model organism *E. coli* K-12, which possesses a two-gene remnant of an ablated flagellar system (Flag-2)[21]. At least one degenerate flagellar system (from *Brucella melitensis*) still has a cryptic role in infection, even though it is no longer capable of mediating flagellar motility — proof that flagellar systems can mediate a useful function other than locomotion[22].

Beyond the common ancestor

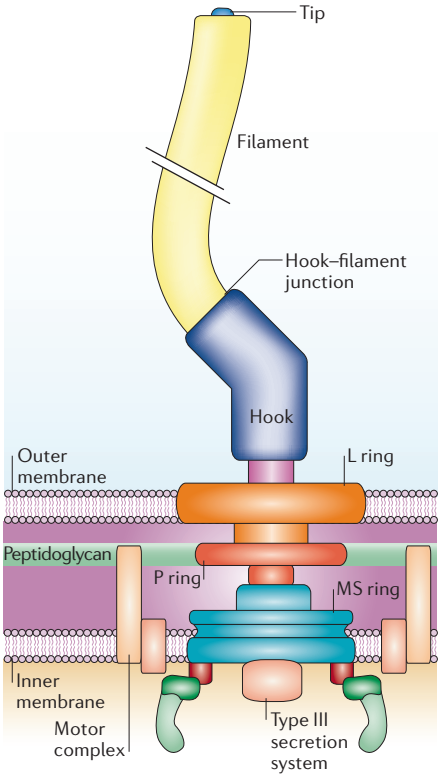
Despite this diversity, it is clear that all (bacterial) flagella share a conserved core set of proteins. Of the forty or so proteins in the standard flagellum of *S.typhimurium* strain LT2 or *E. coli* K-12, only about half seem to be universally necessary (TABLE 1). This reduced flagellum is still a challenge to explain, but if one accepts that all current flagellar systems diverged from their last common ancestor (the ur-flagellum), why stop there? All flagellins show sequence similarity indicative of common ancestry (homology)[15,23]. But then all flagellins also share homology with another component of the flagellar filament, the hook-associated protein 3 (HAP3) or FlgL (as is evident from the application of InterProScan to FlgL from *E. coli*)[15,24]. Therefore, although the ur-flagellum contained flagellin and HAP3, these two proteins must have evolved from a common ancestor in a simpler system that contained only one flagellin-HAP3 homologue. Similarly, six proteins from the rod (FlgB, FlgC, FlgF and FlgG), hook (FlgE) and filament (HAP1/FlgK) show sequence similarities indicative of common ancestry[25]. Therefore, the flagellar rod–hook–filament complex has clearly evolved by multiple rounds of gene duplication and subsequent diversification, starting from just two proteins (a proto-flagellin and a proto-rod/hook protein) that were capable of polymerization into an axial arrangement.

When examining homologies between flagellar and non-flagellar (NF) proteins, it becomes clear that several flagellar subunits share common ancestry with components from other biological systems. For example, at least two regulatory components are homologous to NF proteins: FliK resembles YscP from the Ysc–Yop NF T3SS, and the flagellar σ factor, FliA, shows clear sequence and structural homology to several other NF σ factors[25–28]. Similarly, FlgA, which has a role in the assembly of the P ring, has recently been shown to be homologous to CpaB, a protein involved in type IV pilus assembly[29]. More specifically, both proteins contain a pair of tandem β-clip domains that, it has been hypothesized, might bind to sugars in peptidoglycan[29]. On a similar theme, FlgJ contains a C-terminal amidase 4 domain, which mediates hydrolysis of peptidoglycan and is shared with many other bacterial proteins[30].

What of the flagellar machinery itself? Three modular molecular devices are at the heart of the bacterial flagellum: the rotor-stator that powers flagellar rotation, the chemotaxis apparatus that mediates changes in the direction of motion and the T3SS that mediates export of the axial components of the flagellum. In each module, the apparatus is fashioned from recycled parts that occur elsewhere in nature. Starting with the motor, MotA/MotB are homologous to components of the Ton and Tol systems (ExbB/ExbD and TolQ/TolR). These are systems which also exploit an ion-motive force, not for motility, but instead to drive active transport of substrates across the outer membrane[31]. FliM and FliN share a C-terminal SpoA domain with each other and with components of NF T3SSs (the YscQ-like proteins)[25]. The

Box 1 | **The bacterial flagellum**

The flagellum is the main organelle for motility in bacteria. Despite bearing the same name, bacterial flagella are distinct in form, function and evolution from both archaeal and eukaryotic flagella. The archetypal bacterial flagellum from *Salmonella enterica* serovar Typhimurium (*Salmonella typhimurium*) consists of a basal body, embedded in the cell wall, and two axial structures, the hook and filament, which are joined at the hook–filament junction (see figure). The basal body consists of the MS ring, rod, and L- and P-rings. Components of the axial structures are exported from the cell by the flagellar type III secretion system, which consists of several proteins from the MS ring and a peripheral hexameric ATPase FliI that drives the export process. Exported proteins pass through a central pore in the axial structures — therefore, first the flagellar hook and then the filament grow by the incorporation of new subunits at the distal end of the filamentous structure.



Rotation of bacterial flagella is powered by a proton- or sodium-motive force. The flagellar motor converts electrochemical energy into torque through an interaction between two components: the stator and the rotor. The stator consists of multiple copies of two proteins, MotA and MotB, which assemble into a structure that is associated with the inner membrane and anchored to peptidoglycan, so that it remains stationary. The rotor consists of multiple copies of FliG, which together with FliM and FliN form the C ring, mounted on the cytoplasmic face of the MS ring. Torque is transmitted from the C-ring by the MS ring to the rod (a molecular drive shaft) and from there to the hook (a universal joint) and then on to the helical flagellar filament (a molecular propeller). Rotation of this helical filament converts torque into thrust, conferring motility on the cell. The chemotaxis apparatus (not shown) integrates diverse signals to modulate the behaviour of the motor so as to propel the cell towards nutrients. Several soluble factors are involved in coordinating the assembly of the flagellar apparatus, including the flagellar σ factor FliA and the hook-length control protein, FliK.

An excellent movie illustrating flagellar biosynthesis and structure is available, see Further information. L, lipopolysaccharide; MS, membrane/super membrane; P, peptidoglycan.

chemotaxis system is clearly modular, in that the main components of the chemotaxis apparatus are shared between bacterial and archaeal flagellar systems even though these systems are otherwise completely unrelated (see below)[32,33]. Furthermore, some bacteria, such as *Aquifex aeolicus*, produce conventional bacterial flagella, but lack the equipment for chemotaxis[33]. Finally, many chemotaxis proteins contain domains (for example, response regulator domains), which are found in many other bacterial proteins.

At least nine core flagellar proteins share common ancestry with core components of NF T3SSs. However, the scientific community is divided on the nature of their common ancestor. Some have argued that the NF T3SS was derived from the flagellum, based on the eukaryote-specific function of known NF T3SS, and the apparently limited distribution of known NF T3SSs relative to the broad phylogenetic distribution of flagella[34]. We, and others, have argued that the two systems are sister groups, as indicated by gene phylogenies and arguments based on parsimony, diverging through 'descent with modification' from a common, but simpler, ancestral secretion system[35,36]. Regardless of the conclusion of this debate, the existence of NF T3SSs is 'proof of concept' that a flagellar subsystem can function for purposes other than motility.

There are hints at an even deeper ancestry for parts of this secretion apparatus. FliI, the ATPase that powers flagellar protein export, shows unequivocal homology to the catalytic subunits from F-type and V-type ATPases[37]. More recently, we have uncovered evidence of homology between FliH, a flagellar protein that interacts with FliI, and second-stalk components of the F-type and V-type ATPases that interact with the catalytic subunits[38]. These sequence homologies allude to the existence of ancient structural and functional similarities between these two types of molecular rotary motor, which we predict will become ever more apparent as work on these systems progresses. Given these homologies, it is also interesting to speculate whether the T3SS ATPase initially powered flagellar rotation, before the current more efficient motor components were recruited.

Many paths to motility

Although the evolution by random mutation and natural selection of something as complex as a contemporary bacterial flagellum might, in retrospect, seem highly improbable, it is important to appreciate that probabilities should be assessed by looking forward not back[2]. For example, from studies on protein design it is clear that creating proteins from scratch that, like flagellin, self-assemble into filaments is not very difficult[39,40]. Furthermore, it is clear that there are many other filamentous surface structures in bacteria that show no apparent evolutionary relationship to bacterial flagella[41,42]. In other words, there are plenty of potential starting points for the evolution of a molecular propeller. Evolution of something like the flagellar filament is therefore far less surprising than it might at first seem. In fact, microorganisms have adopted other routes to motility besides the bacterial flagellum[43]. Most strikingly, although archaeal flagella superficially resemble bacterial flagella, in that they too are rotary structures driven by a proton gradient, they are fundamentally distinct from their bacterial counterparts in terms of protein composition and assembly.

Intermediate forms

What about intermediate forms between bacterial flagella and other biological entities? Darwin encountered a similar argument about gaps in the fossil record, and in response he pointed out how improbable fossilization was, so that little of any extinct biosphere could ever be expected to appear in the fossil record[14]. Although fossils are of no use in reconstructing flagellar evolution, similar arguments might be made at the molecular level. Despite a decade of bacterial genome sequencing, we have scarcely begun to sample the molecular diversity of the biosphere. Yet even with the scant coverage of genome sequence data to date, several curiosities have already been revealed. For example, there is growing evidence that flagellin and the flagellar filament are homologous to the NF T3SS protein EspA and the EspA filament, respectively[35,44–48]. The EspA filament therefore provides a model for how the ancestral flagellar filament might have functioned for purposes other than locomotion (adhesion or targeted protein secretion). Furthermore, the EspA protein from *E. coli* initially seemed to be one of a kind. However, thanks to genome sequencing, related proteins have been identified in several bacteria occupying diverse niches, including: *S. typhimurium*, *Edwardsiella ictaluri*, *Shewanella baltica*, *Chromobacterium violaceum*, *Yersinia frederiksenii*, *Yersinia bercovieri* and *Sodalis glossinidius*. In addition, proteins that resemble flagellar components but that are encoded in the genomes of bacteria that do not engage in flagellar motility have also been identified. The first example of these potential 'missing links' came from the chlamydias[49]. More recently, flagellar-related genes have been detected in the genome of the soil bacterium *Myxococcus xanthus*, which uses gliding rather than flagellar motility[35]. It seems likely that other examples of potential evolutionary intermediaries will be found as we sequence an increasing proportion of the biosphere.

Box 2 | **Of forelimbs and flagella**

ID advocates say that their position is supported by discontinuities between the flagellum and the rest of the biological world, just as a designed entity like a watch differs from an undesigned entity, such as a stone. In support of this line of reasoning, Scott Minnich in his expert witness report claimed that "the other thirty proteins in the flagellar motor (that are not present in the type III secretion system) are unique to the motor and are not found in any other living system." As our discussion shows, this is not true. Instead, we have detected sequence homologies linking flagellar components to the rest of the biological universe (TABLE 1). Furthermore, a fundamental evolutionary insight underlies the very principle of homology, whether in anatomy or amino-acid sequence: there are organizational similarities in biology that cannot be explained by 'design' for a particular function. Richard Owen first made this argument with tetrapod forelimbs, pointing out that they shared fundamental organizational similarities despite being used for digging, walking, swimming, flight, and in the case of humans, tool use. Darwin reiterated the argument, this time in defence of evolution, in *The Origin of Species*:

"What can be more curious than that the hand of a man, formed for grasping, that of a mole for digging, the leg of the horse, the paddle of the porpoise, and the wing of the bat, should all be constructed on the same pattern, and should include the same bones, in the same relative positions?... Nothing can be more hopeless than to attempt to explain this similarity of pattern in members of the same class, by utility or by the doctrine of final causes. The hopelessness of the attempt has been expressly admitted by Owen in his most interesting work on the 'Nature of Limbs.' On the ordinary view of the independent creation of each being, we can only say that so it is; that it has pleased the Creator to construct all the animals and plants in each great class on a uniform plan; but this is not a scientific explanation."

Table 1 | **Homologies of flagellar proteins**

Protein	Location	Function	Indispensable?	Homologies*	Refs
FlgA	P ring	Chaperone?	Absent from Gram-positive bacteria	CpaB‡	25, 29
FlgBCFG	Rod	Transmission shaft	Yes	FlgBCEFGK§	25
FlgD	Hook	Hook cap	Yes		25
FlgE	Hook	Universal joint	Yes	FlgBCEFGK	25
FlgH	L ring	Bushing	Absent from Gram-positive bacteria	None yet known	25
FlgI	P ring	Bushing	Absent from Gram-positive bacteria	None yet known	25
FlgJ	Rod	Rod cap; muramidase	FlgJ N-terminal domain absent from some systems	None yet known	25
FlgK	Hook–filament junction	Hook-associated protein 1	Yes	FlgBCEFGK§	25
FlgL	Hook–filament junction	Hook-associated protein 3	Yes	FliC§	25
FlgM	Cytoplasm and exterior	Anti-σ factor	Absent from *Caulobacter*	None yet known	25
FlgN	Cytoplasm	Chaperone	Undetectable in some systems	None yet known	25
FlhA	T3SS apparatus	Protein export	Yes	LcrD/YscV‖	25
FlhB	T3SS apparatus	Protein export	Yes	YscU‖	25
FlhDC	Cytoplasm	Transcriptional regulator	Absent from many systems	Other activators‡	25
FlhE	Unknown	Unknown	Mutant retains full motility		25
FliA	Cytoplasm	σ factor	Absent from *Caulobacter*	RpoD, RpoH, RpoS‖	25
FliB	Cytoplasm	N-methylase	Absent from *Escherichia coli*		25
FliC	Filament	Flagellin	Yes	FlgL§, EspA¶	25, 78
FliD	Filament	Filament cap; hook-associated protein 2	Absent from *Caulobacter*	None yet known	25
FliE	Rod/basal body	MS ring–rod junction	Yes	None yet known	25
FliF	T3SS apparatus	Protein export	Yes	YscJ§	25
FliG	Peripheral	Motor	Yes	MgtE¶	25
FliH	T3SS apparatus	Regulates FliI	Mutant retains some motility	YscL*, AtpFH¶	38, 79
FliI	T3SS apparatus	ATPase for protein export	Yes	YscN‖, AtpD‖, Rho‖	38
FliJ	Cytoplasm	Chaperone	Undetectable in some systems	YscO¶	25
FliK	Hook/basal body	Controls hook length	Yes	YscP¶	25
FliL	Basal body	Unknown	Mutant retains full motility	None yet known	80
FliM	T3SS apparatus	Protein export	Yes	FliN‡, YscQ‡	25
FliN	T3SS apparatus	Protein export	Yes	FliM‡, YscQ‡	25
FliO	T3SS apparatus	Protein export	Undetectable in some systems	None	25
FliP	T3SS apparatus	Protein export	Yes	YscR‖	25
FliQ	T3SS apparatus	Protein export	Yes	YscS‖	25
FliR	T3SS apparatus	Protein export	Yes	YscT‖	25
FliS	Cytoplasm	FliC chaperone	Absent from *Caulobacter*	None yet known	25
FliT	Cytoplasm	FliD chaperone	Absent from many systems	None yet known	25
FliZ	Cytoplasm	Regulator	Absent from many systems	None yet known	25
MotA	Inner membrane	Motor	Yes	ExbB‡, TolQ‡	25
MotB	Inner membrane	Motor	Yes	ExbD‡, TolR‡, OmpA‡	25

*Homologies (as evidenced by expected values <1e05) can be confirmed by retrieving the relevant flagellar protein sequences for *Escherichia coli* K-12 or *Salmonella enterica* Typhimurium LT2 and carrying out the following: ‡performing multiple PSI-BLAST iterations at the NCBI site under default conditions; §performing multiple iterations of PSI–BLAST at the NCBI site under default conditions, except for adjusting the threshold for inclusion to 0.05 and restricting the taxonomic scope to Enterobacteriaceae, starting with the protein sequences for FlgB, FlgL; ‖performing a BLASTp search at the NCBI site under default conditions. ¶These similarities fail to achieve unequivocal significance using BLAST/PSI–BLAST under any of the above conditions, but are supported by other structural or functional considerations. T3SS, type III secretion system.

Towards a plausible evolutionary model

From the above discussions of sequence homologies and modularity, it is clear that designing an evolutionary model to account for the origin of the ancestral flagellum requires no great conceptual leap. Instead, one can envisage the ur-flagellum arising from mergers between several modular subsystems: a secretion system built from proteins accreted around an ancient ATPase, a filament built from variants of two initial proteins, a motor built from an ion channel and a chemotaxis apparatus built from pre-existing regulatory domains (FIG. 1). As we have seen, each of these function in a modular fashion and share ancestry with simpler systems — thereby answering the question 'what use is half a flagellum?' Furthermore, it is not hard to envisage how an ancestral crude and inefficient flagellum, if it conferred any motility at all, could function as the starting material for natural selection to fashion today's slicker flagellar apparatus.

However, one could still question how, from such bricolage, natural selection could lock on to an evolutionary trajectory leading to an organelle of motility in the first place, when none of the components alone confer the organism with a selective advantage relevant to motility. The key missing concept here is that of exaptation, in which the function currently performed by a biological system is different from the function performed while the adaptation evolved under earlier pressures of natural selection[50]. For example, a bird's feathers might have originally arisen in the context of selection for, say, heat control, and only later have been used to assist with flight[51,52]. Under this argument, a number of slight but decisive functional shifts occurred in the evolution of the flagellum, the most recent of which was probably a shift from an organelle of adhesion or targeted secretion, such as the EspA filament, to a curved structure capable of generating a propulsive force.

An experimental research programme?

The famous evolutionist Theodosius Dobzhansky once stated that "nothing in biology makes sense except in the light of evolution"[53]. In recent years, flagellar biologists have made astonishing progress in understanding the structure, function and regulation of bacterial flagella[54–64]. In addition, they have moved beyond the *S. typhimurium* paradigm to embrace a richly diverse set of systems. For example, a recent review on flagellar regulation[65] highlighted the assorted regulatory pathways governing flagellar function in the peritrichous flagella of enteric bacteria, the polar flagellar systems of α-, γ- and ε-proteobacteria, the lateral flagellar system of *Vibrio parahaemolyticus*, the endoflagella of spirochaetes and the flagella of Gram-positive bacteria.

However, the flagellar research community has scarcely begun to consider how these systems have evolved. This neglect probably stems from a reluctance to engage in the 'armchair speculation' inherent in building evolutionary models, and from a desire to determine how a system works before wondering how it got to be that way. However, there are several good reasons for adopting an evolutionary approach to flagellar biology. Assignments of homology can provide insights into function, and can provide a framework for interpreting the sequence data in the post-genomic era. The abundance of these data indicates that current studies are looking at the 'tip of an iceberg'. Recently, genome sequencing revealed that *Desulfotalea psychrophila*, a sulphate-reducing bacterium from permanently cold Arctic sediments[66], has the largest of all known flagellin genes, but without a 'bigger-picture' view of flagellar biology, we have no idea why. Furthermore, an evolutionary comparative approach fits in perfectly with the current zeitgeist, with its emphasis on evolutionary systems biology[67].

Notwithstanding the good scientific reasons for new forays in this direction, the lack of a scientific literature on flagellar evolution also has another undesirable consequence — it leaves open the suspicion among members of the public that maybe there is some mystery here, that maybe the ID proponents do have a point. Although all experts in this field agree that there is nothing to these claims, as Wilkins has recently pointed out[68], in these politically charged times, it is no longer enough to say, for example, that bacterial flagella evolved and that is that. Instead, scientific experts have to engage with a sceptical public.

Scott Minnich speculated in his testimony that studies on flagellar evolution need not be restricted to sequence analysis or theoretical models, but that instead this topic could become the subject of laboratory-based experimental studies. But obviously, one cannot model millions of years of evolution in a few weeks or months. So how might such studies be conducted? One option might be to look back in time. It is feasible to use phylogenetic analyses to reconstruct plausible ancestral sequences of modern-day proteins, and then synthesize and investigate these ancestral proteins. Proof of principle for this approach has already been demonstrated on several NF proteins[69–75]. Similar studies could recreate plausible ancestors

Glossary

β-clip domain
A fold found in a diverse group of protein domains typified by the presence of two characteristic waist-like constrictions, flanking a central extended region. The flagellar P-ring protein FlgA and type IV pilus assembly protein CpaB are two examples of β-clip-domain-containing proteins.

Chemotaxis
A behavioural response by bacteria whereby a bacterial cell senses a chemical gradient and moves towards or away from the chemical stimulus.

Essentialism
Also referred to as typology. The idea that a specific kind of entity can be defined by an invariant essence. A triangle illustrates essentialism: all triangles have the same fundamental characteristics and are sharply delimited against quadrangles or any other geometric figures. An intermediate between a triangle and a quadrangle is inconceivable. Typological thinking is however unable to accommodate the profligate variation that occurs in biology.

Establishment clause
A clause from the First Amendment to the American Constitution that states that: 'Congress shall make no law respecting an establishment of religion'. This is now interpreted to forbid any state funding of religious education in the United States.

Intelligent design
(ID). The concept that some aspects of the natural universe are better explained by an intelligent cause rather than by an undirected process such as natural selection.

Irreducible complexity
The notion that some biological systems are so complex that they could not function if they were any simpler, and so could not have been formed by successive additions to a precursor system with the same functionality.

Occam's razor
The principle that the explanation of any phenomenon should make as few assumptions as possible.

Proton-motive force
Storage of energy as a combination of a proton and voltage gradient across the bacterial inner membrane. The proton-motive force is exploited by the membrane-associated F-type ATPase to generate ATP, and by the flagellar motor to generate torque. In some bacteria, an analogous sodium-motive force drives flagellar rotation.

SpoA domain
A β-sheet domain found at the C terminus of flagellar proteins FliM and FliN and non-flagellar T3SS proteins such as YscQ and HrcQb.

for various flagellar components (for example, the common ancestor of flagellins and HAP3 proteins). These proteins could then be reproduced in the laboratory in order to examine their properties (for example, how well they self-assemble into filaments and what those filaments look like). An alternative, more radical, option would be to model flagellar evolution prospectively, for example, by creating random or minimally constrained libraries and then iteratively selecting proteins that assemble into ever more sophisticated artificial analogues of the flagellar filament. Another experimental option might be to investigate the environmental conditions that favour or disfavour bacterial motility. The fundamental physics involved (diffusion due to Brownian motion) is mathematically tractable, and has already been used to predict, for example, that powered motility is useless in very small bacteria[76,77].

The final word

Like Darwin, we have found that careful attention to homology, analogy and diversity yields substantial insights into the origin of even the most complex systems. We close with a quotation from the closing chapter of *The Origin of Species* that applies as well to a bacterial flagellum as to any other evolved entity:

"When we no longer look at an organic being as a savage looks at a ship, as something wholly beyond his comprehension; when we regard every production of nature as one which has had a long history; when we contemplate every complex structure and instinct as the summing up of many contrivances, each useful to the possessor, in the same way as any great mechanical invention is the summing up of the labour, the experience, the reason, and even the blunders of numerous workmen; when we thus view each organic being, how far more interesting — I speak from experience — does the study of natural history become!"[14].

Mark J. Pallen is at the Division of Immunity & Infection, Medical School, University of Birmingham, Birmingham, B15 2TT UK.

Nicholas J. Matzke is with the National Center for Science Education, Oakland, California 94609–2509 USA.

Correspondence to M.J.P.
e-mail: m.pallen@bham.ac.uk

doi: 10.1038/nrmicro1493
Published online 5 September 2006

1. Musgrave, I. in *Why Intelligent Design Fails: a Scientific Critique of the New Creationism* (eds Young, M. & Edis, T.) 72–84 (Rutgers University Press, Piscataway USA, 2004).
2. Miller, K. R. in *Debating Design: from Darwin to DNA* (eds Dembski, W. & Ruse, M.) 81–97 (Cambridge University Press, New York, 2004).
3. Mayr, E. Darwin's influence on modern thought. *Sci. Am.* **283**, 78–83 (2000).
4. Kubori, T. *et al.* Purification and characterization of the flagellar hook-basal body complex of *Bacillus subtilis*. *Mol. Microbiol.* **24**, 399–410 (1997).
5. Li, C., Motaleb, A., Sal, M., Goldstein, S. F. & Charon, N. W. Spirochete periplasmic flagella and motility. *J. Mol. Microbiol. Biotechnol.* **2**, 345–354 (2000).
6. McCarter, L. L. Dual flagellar systems enable motility under different circumstances. *J. Mol. Microbiol. Biotechnol.* **7**, 18–29 (2004).
7. Kita-Tsukamoto, K., Wada, M., Yao, K., Nishino, T. & Kogure, K. Flagellar motors of marine bacteria *Halomonas* are driven by both protons and sodium ions. *Can. J. Microbiol.* **50**, 369–374 (2004).
8. Armitage, J. P. & Macnab, R. M. Unidirectional, intermittent rotation of the flagellum of *Rhodobacter sphaeroides*. *J. Bacteriol.* **169**, 514–518 (1987).
9. Attmannspacher, U., Scharf, B. & Schmitt, R. Control of speed modulation (chemokinesis) in the unidirectional rotary motor of *Sinorhizobium meliloti*. *Mol. Microbiol.* **56**, 708–718 (2005).
10. Shibata, S., Alam, M. & Aizawa, S. Flagellar filaments of the deep-sea bacteria *Idiomarina ioihiensis* belong to a family different from those of *Salmonella typhimurium*. *J. Mol. Biol.* **352**, 510–516 (2005).
11. Burnens, A. P. *et al.* The flagellin N-methylase gene *fliB* and an adjacent serovar-specific IS*200* element in *Salmonella typhimurium*. *Microbiology* **143**, 1539–1547 (1997).
12. Logan, S. M. Flagellar glycosylation — a new component of the motility repertoire? *Microbiology* **152**, 1249–1262 (2006).
13. Read, T. D. *et al.* Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res.* **28**, 1397–1406 (2000).
14. Darwin, C. *The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (John Murray, London, 1859).
15. Beatson, S. A., Minamino, T. & Pallen, M. J. Variation in bacterial flagellins: from sequence to structure. *Trends Microbiol.* **14**, 151–155 (2006).
16. Ely, B., Ely, T. W., Crymes, W. B. Jr & Minnich, S. A. A family of six flagellin genes contributes to the *Caulobacter crescentus* flagellar filament. *J. Bacteriol.* **182**, 5001–5004 (2000).
17. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
18. Halling, S. M. On the presence and organization of open reading frames of the nonmotile pathogen *Brucella abortus* similar to class II, III, and IV flagellar genes and to LcrD virulence superfamily. *Microb. Comp. Genomics* **3**, 21–29 (1998).
19. Al Mamun, A. A., Tominaga, A. & Enomoto, M. Cloning and characterization of the region III flagellar operons of the four *Shigella* subgroups: genetic defects that cause loss of flagella of *Shigella boydii* and *Shigella sonnei*. *J. Bacteriol.* **179**, 4493–4500 (1997).
20. Monday, S. R., Minnich, S. A. & Feng, P. C. A 12-base-pair deletion in the flagellar master control gene *flhC* causes nonmotility of the pathogenic German sorbitol-fermenting *Escherichia coli* O157: H- strains. *J. Bacteriol.* **186**, 2319–2327 (2004).
21. Ren, C. P., Beatson, S. A., Parkhill, J. & Pallen, M. J. The Flag-2 locus, an ancestral gene cluster, is potentially associated with a novel flagellar system from *Escherichia coli*. *J. Bacteriol.* **187**, 1430–1440 (2005).
22. Fretin, D. *et al.* The sheathed flagellum of *Brucella melitensis* is involved in persistence in a murine model of infection. *Cell. Microbiol.* **7**, 687–698 (2005).
23. Webber, C. & Ponting, C. P. Genes and homology. *Curr. Biol.* **14**, R332–R333 (2004).
24. Quevillon, E. *et al.* InterProScan: protein domains identifier. *Nucleic Acids Res.* **33**, W116–W120 (2005).
25. Pallen, M. J., Penn, C. W. & Chaudhuri, R. R. Bacterial flagellar diversity in the post-genomic era. *Trends Microbiol.* **13**, 143–149 (2005).
26. Agrain, C. *et al.* Characterization of a Type III secretion substrate specificity switch (T3S4) domain in YscP from *Yersinia enterocolitica*. *Mol. Microbiol.* **56**, 54–67 (2005).
27. Ohnishi, K., Kutsukake, K., Suzuki, H. & Iino, T. Gene *fliA* encodes an alternative σ factor specific for flagellar operons in *Salmonella typhimurium*. *Mol. Gen. Genet.* **221**, 139–147 (1990).
28. Sorenson, M. K., Ray, S. S. & Darst, S. A. Crystal structure of the flagellar σ/anti-σ complex σ^{28}/FlgM reveals an intact σ factor in an inactive conformation. *Mol. Cell* **14**, 127–138 (2004).
29. Iyer, L. M. & Aravind, L. The emergence of catalytic and structural diversity within the β-clip fold. *Proteins* **55**, 977–991 (2004).
30. Nambu, T., Minamino, T., Macnab, R. M. & Kutsukake, K. Peptidoglycan-hydrolyzing activity of the FlgJ protein, essential for flagellar rod formation in *Salmonella typhimurium*. *J. Bacteriol.* **181**, 1555–1561 (1999).
31. Zhai, Y. F., Heijne, W. & Saier, M. H. Jr. Molecular modeling of the bacterial outer membrane receptor energizer, ExbBD/TonB, based on homology with the flagellar motor, MotAB. *Biochim. Biophys. Acta* **1614**, 201–210 (2003).
32. Szurmant, H. & Ordal, G. W. Diversity in chemotaxis mechanisms among the bacteria and archaea. *Microbiol. Mol. Biol. Rev.* **68**, 301–319 (2004).
33. Faguy, D. M. & Jarrell, K. F. A twisted tale: the origin and evolution of motility and chemotaxis in prokaryotes. *Microbiology* **145**, 279–281 (1999).
34. Nguyen, L., Paulsen, I. T., Tchieu, J., Hueck, C. J. & Saier, M. H. Jr. Phylogenetic analyses of the constituents of Type III protein secretion systems. *J. Mol. Microbiol. Biotechnol.* **2**, 125–144 (2000).
35. Pallen, M. J., Beatson, S. A. & Bailey, C. M. Bioinformatics, genomics and evolution of non-flagellar type-III secretion systems: a Darwinian perspective. *FEMS Microbiol. Rev.* **29**, 201–229 (2005).
36. Gophna, U., Ron, E. Z. & Graur, D. Bacterial type III secretion systems are ancient and evolved by multiple horizontal-transfer events. *Gene* **312**, 151–163 (2003).
37. Vogler, A. P., Homma, M., Irikura, V. M. & Macnab, R. M. *Salmonella typhimurium* mutants defective in flagellar filament regrowth and sequence similarity of FliI to F0F1, vacuolar, and archaebacterial ATPase subunits. *J. Bacteriol.* **173**, 3564–3572 (1991).
38. Pallen, M. J., Bailey, C. M. & Beatson, S. A. Evolutionary links between FliH/YscL-like proteins from bacterial type III secretion systems and second-stalk components of the F0F1 and vacuolar ATPases. *Protein Sci.* **15**, 935–941 (2006).
39. Woolfson, D. N. The design of coiled-coil structures and assemblies. *Adv. Protein Chem.* **70**, 79–112 (2005).
40. Wagner, D. E. *et al.* Toward the development of peptide nanofilaments and nanoropes as smart materials. *Proc. Natl Acad. Sci. USA* **102**, 12656–12661 (2005).
41. Fernandez, L. A. & Berenguer, J. Secretion and assembly of regular surface structures in Gram-negative bacteria. *FEMS Microbiol. Rev.* **24**, 21–44 (2000).
42. Ton-That, H. & Schneewind, O. Assembly of pili in Gram-positive bacteria. *Trends Microbiol.* **12**, 228–234 (2004).
43. Bardy, S. L., Ng, S. Y. & Jarrell, K. F. Prokaryotic motility structures. *Microbiology* **149**, 295–304 (2003).
44. Knutton, S. *et al.* A novel EspA-associated surface organelle of enteropathogenic *Escherichia coli* involved in protein translocation into epithelial cells. *EMBO J.* **17**, 2166–2176 (1998).
45. Delahay, R. M. *et al.* The coiled-coil domain of EspA is essential for the assembly of the type III secretion translocon on the surface of enteropathogenic *Escherichia coli*. *J. Biol. Chem.* **274**, 35969–35974 (1999).
46. Daniell, S. J. *et al.* 3D structure of EspA filaments from enteropathogenic *Escherichia coli*. *Mol. Microbiol.* **49**, 301–308 (2003).

47. Crepin, V. F., Shaw, R., Abe, C. M., Knutton, S. & Frankel, G. Polarity of enteropathogenic *Escherichia coli* EspA filament assembly and protein secretion. *J. Bacteriol.* **187**, 2881–2889 (2005).
48. Yip, C. K., Finlay, B. B. & Strynadka, N. C. Structural characterization of a type III secretion system filament protein in complex with its chaperone. *Nature Struct. Mol. Biol.* **12**, 75–81 (2005).
49. Kim, J. F. Revisiting the chlamydial type III protein secretion system: clues to the origin of type III protein secretion. *Trends Genet.* **17**, 65–69 (2001).
50. Gould, S. J. & Vrba, E. S. Exaptation — a missing term in the science of form. *Paleobiology* **8**, 4–15 (1982).
51. Regal, P. J. The evolutionary origin of feathers. *Q. Rev. Biol.* **50**, 35–66 (1975).
52. Norell, M. *et al.* Palaeontology: 'modern' feathers on a non-avian dinosaur. *Nature* **416**, 36–37 (2002).
53. Dobzhansky, T. Nothing in biology makes sense except in the light of evolution. *Am. Biol. Teach.* **35**, 125–129 (1973).
54. Macnab, R. M. Type III flagellar protein export and flagellar assembly. *Biochim. Biophys. Acta* **1694**, 207–217 (2004).
55. Minamino, T. & Namba, K. Self-assembly and type III protein export of the bacterial flagellum. *J. Mol. Microbiol. Biotechnol.* **7**, 5–17 (2004).
56. Suzuki, H., Yonekura, K. & Namba, K. Structure of the rotor of the bacterial flagellar motor revealed by electron cryomicroscopy and single-particle image analysis. *J. Mol. Biol.* **337**, 105–113 (2004).
57. Saijo-Hamano, Y., Minamino, T., Macnab, R. M. & Namba, K. Structural and functional analysis of the C-terminal cytoplasmic domain of FlhA, an integral membrane component of the type III flagellar protein export apparatus in *Salmonella*. *J. Mol. Biol.* **343**, 457–466 (2004).
58. Samatey, F. A. *et al.* Structure of the bacterial flagellar protofilament and implications for a switch for supercoiling. *Nature* **410**, 331–337 (2001).
59. Yonekura, K., Maki-Yonekura, S. & Namba, K. Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature* **424**, 643–650 (2003).
60. Journet, L., Hughes, K. T. & Cornelis, G. R. Type III secretion: a secretory pathway serving both motility and virulence. *Mol. Membr. Biol.* **22**, 41–50 (2005).
61. Aldridge, P. & Hughes, K. T. Regulation of flagellar assembly. *Curr. Opin. Microbiol.* **5**, 160–165 (2002).
62. Sowa, Y. *et al.* Direct observation of steps in rotation of the bacterial flagellar motor. *Nature* **437**, 916–919 (2005).
63. Brown, P. N., Hill, C. P. & Blair, D. F. Crystal structure of the middle and C-terminal domains of the flagellar rotor protein FliG. *EMBO J.* **21**, 3225–3234 (2002).
64. Blair, D. F. Flagellar movement driven by proton translocation. *FEBS Lett.* **545**, 86–95 (2003).
65. McCarter, L. L. Regulation of flagella. *Curr. Opin. Microbiol.* **9**, 180–186 (2006).
66. Rabus, R. *et al.* The genome of *Desulfotalea psychrophila*, a sulfate-reducing bacterium from permanently cold Arctic sediments. *Environ. Microbiol.* **6**, 887–902 (2004).
67. Medina, M. Genomes, phylogeny, and evolutionary systems biology. *Proc. Natl Acad. Sci. USA* **102**, (Suppl. 1) 6630–6635 (2005).
68. Wilkins, A. S. 'Intelligent design' as both problem and symptom. *Bioessays* **28**, 327–329 (2006).
69. Field, S. F., Bulina, M. Y., Kelmanson, I. V., Bielawski, J. P. & Matz, M. V. Adaptive evolution of multicolored fluorescent proteins in reef-building corals. *J. Mol. Evol.* **62**, 332–339 (2006).
70. Chang, B. S., Ugalde, J. A. & Matz, M. V. Applications of ancestral protein reconstruction in understanding protein function: GFP-like proteins. *Methods Enzymol* **395**, 652–670 (2005).
71. Wouters, M. A., Liu, K., Riek, P. & Husain, A. A despecialization step underlying evolution of a family of serine proteases. *Mol. Cell* **12**, 343–354 (2003).
72. Ugalde, J. A., Chang, B. S. & Matz, M. V. Evolution of coral pigments recreated. *Science* **305**, 1433 (2004).
73. Chang, B. S. & Donoghue, M. J. Recreating ancestral proteins. *Trends Ecol. Evol.* **15**, 109–114 (2000).
74. Chang, B. S., Kazmi, M. A. & Sakmar, T. P. Synthetic gene technology: applications to ancestral gene reconstruction and structure-function studies of receptors. *Meth. Enzymol.* **343**, 274–294 (2002).
75. Chang, B. S., Jonsson, K., Kazmi, M. A., Donoghue, M. J. & Sakmar, T. P. Recreating a functional ancestral archosaur visual pigment. *Mol. Biol. Evol.* **19**, 1483–1489 (2002).
76. Dusenbery, D. B. Fitness landscapes for effects of shape on chemotaxis and other behaviors of bacteria. *J. Bacteriol.* **180**, 5978–5983 (1998).
77. Dusenbery, D. B. Minimum size limit for useful locomotion by free-swimming microbes. *Proc. Natl Acad. Sci. USA* **94**, 10949–10954 (1997).
78. Pallen, M. J., Beatson, S. A. & Bailey, C. M. Bioinformatics analysis of the locus for enterocyte effacement provides novel insights into type-III secretion. *BMC Microbiol.* **5**, 9 (2005).
79. Minamino, T., Gonzalez-Pedrajo, B., Kihara, M., Namba, K. & Macnab, R. M. The ATPase FliI can interact with the type III flagellar protein export apparatus in the absence of its regulator, FliH. *J. Bacteriol.* **185**, 3983–3988 (2003).
80. Raha, M., Sockett, H. & Macnab, R. M. Characterization of the *fliL* gene in the flagellar regulon of *Escherichia coli* and *Salmonella typhimurium*. *J. Bacteriol.* **176**, 2308–2311 (1994).

Acknowledgements
We thank D. Blair, P. Aldridge and R. Berry for critical comments on this manuscript.

Competing interests statement
The authors declare competing financial interests: see web version for details.

DATABASES
The following terms in this article are linked online to:
Entrez Genome Project: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeprj
Aquifex aeolicus | *Brucella melitensis* | *Chromobacterium violaceum* | *Desulfotalea psychrophila* | *Edwardsiella ictaluri* | *Escherichia coli* | *Myxococcus xanthus* | *Rhodobacter sphaeroides* | *Salmonella typhimurium* | *Shewanella baltica* | *Sinorhizobium meliloti* | *Sodalis glossinidius* | *Vibrio parahaemolyticus* | *Yersinia bercovieri* | *Yersinia frederiksenii*

FURTHER INFORMATION
Mark Pallen's homepage: http://www.infection.bham.ac.uk/BPAG/staff/mpallen.html
Evolution in (Brownian) space: a model for the origin of the bacterial flagellum: http://www.talkdesign.org/faqs/flagellum.html
Kitzmiller versus Dover trial information: http://www2.ncseweb.org/wp/?page_id=5
Ken Miller's 'the flagellum unspun': http://www.millerandlevine.com/km/evol/design2/article.html
Nanotechnology Researchers Network Center of Japan: http://www.nanonet.go.jp/english/mailmag/2004/011a.html
The Minnich and Miller expert witness reports: http://www2.ncseweb.org/kvd/experts/minnich.pdf http://www2.ncseweb.org/kvd/experts/miller.pdf
Talk Reason: http://www.talkreason.org
The Pandas Thumb: http://www.pandasthumb.org
The Talk. Origins Archive: http://www.talkorigins.org
Access to this links box is available online.

INNOVATION

Bioinformatics-assisted anti-HIV therapy

Thomas Lengauer and Tobias Sing

Abstract | Highly active antiretroviral therapy (HAART), in which three or more drugs are given in combination, has substantially improved the clinical management of HIV-1 infection. Still, the emergence of drug-resistant variants eventually leads to therapy failure in most patients. In such a scenario, the high diversity of resistance-associated mutational patterns complicates the choice of an optimal follow-up regimen. To support physicians in this task, a range of bioinformatics tools for predicting drug resistance or response to combination therapy from the viral genotype have been developed. With several free and commercial software services available, computational advice is rapidly gaining acceptance as an important element of rational decision-making in the treatment of HIV infection.

Since its recognition in 1981, AIDS has killed more than 25 million people, making it one of the most destructive epidemics in the history of mankind. Despite intensive public health efforts, the number of people living with HIV has reached a new peak of over 40 million in 2005, with 3 million deaths and 5 million new infections in the same year[1]. Modern combination therapy can substantially delay disease progression, prolong survival and maintain quality of life, but a cure for HIV infection remains out of reach. Therefore, research focuses not only on the search for novel drugs, but also on exploiting the currently available drug collection to the best possible effect using personalized therapy administration. The main obstacle to ultimate treatment success is the ability of the virus to rapidly acquire mutations that confer resistance to specific drugs (FIG. 1). To identify the genomic make-up of the viral population at treatment failure, relevant portions of the HIV genome are now routinely sequenced in many countries. These efforts have led to the discovery of an ever-increasing number