MOLECULAR BIOLOGY AND EVOLUTION

Rapid Evolution and the Importance of Recombination to the Gastroenteric Pathogen *Campylobacter jejuni*

Daniel J. Wilson,[1] *Edith Gabriel,*†[2] *Andrew J.H. Leatherbarrow,*‡ *John Cheesbrough,*§ *Steven Gee,*§ *Eric Bolton,*|| *Andrew Fox,*§|| *C. Anthony Hart,*¶[3] *Peter J. Diggle,*† *and Paul Fearnhead**

*Department of Maths and Statistics, Lancaster University, Lancaster, United Kingdom; †Department of Medicine, Lancaster University, Lancaster, United Kingdom; ‡Faculty of Veterinary Science, University of Liverpool, Leahurst, Neston, United Kingdom; §Preston Microbiology Services, Royal Preston Hospital, Lancashire Teaching Hospitals NHS Foundation Trust, Preston, United Kingdom; ||Manchester Medical Microbiology Partnership, P.O. Box 209, Clinical Sciences Building, Manchester Royal Infirmary, Manchester, United Kingdom; and ¶Division of Medical Microbiology, School of Infection and Host Defence, University of Liverpool, Liverpool, United Kingdom

Responsible for the majority of bacterial gastroenteritis in the developed world, *Campylobacter jejuni* is a pervasive pathogen of humans and animals, but its evolution is obscure. In this paper, we exploit contemporary genetic diversity and empirical evidence to piece together the evolutionary history of *C. jejuni* and quantify its evolutionary potential. Our combined population genetics–phylogenetics approach reveals a surprising picture. *Campylobacter jejuni* is a rapidly evolving species, subject to intense purifying selection that purges 60% of novel variation, but possessing a massive evolutionary potential. The low mutation rate is offset by a large effective population size so that a mutation at any site can occur somewhere in the population within the space of a week. Recombination has a fundamental role, generating diversity at twice the rate of de novo mutation, and facilitating gene flow between *C. jejuni* and its sister species *Campylobacter coli*. We attempt to calibrate the rate of molecular evolution in *C. jejuni* based solely on within-species variation. The rates we obtain are up to 1,000 times faster than conventional estimates, placing the *C. jejuni–C. coli* split at the time of the Neolithic revolution. We weigh the plausibility of such recent bacterial evolution against alternative explanations and discuss the evidence required to settle the issue.

Introduction

The World Health Organization expects that every year 1% of the population of developed nations will suffer campylobacteriosis (Humphrey et al. 2007), a diarrheal disease that can lead to serious sequelae such as Guillain–Barré syndrome and reactive arthritis (Zia et al. 2003). *Campylobacter jejuni* is the principal bacterial agent responsible for gastroenteritis, ahead of *Salmonella*, *Escherichia coli*, *Clostridium*, and *Listeria* combined (Adak et al. 2005). For such a common pathogen, surprisingly little is known about its evolution. What we do know is that *Campylobacter* species are zoonotic pathogens that colonize the gut of a wide variety of birds and mammals. *Campylobacter jejuni*, the species responsible for 90% of human disease, is found commonly in cattle, sheep, pigs, poultry, wild birds, rabbits, other wild mammals, household pets, molluscs, sewage, and in natural water sources such as rivers and the coast (Jones 2001; Humphrey et al. 2007). The *C. jejuni* gene pools in different host species are largely overlapping (McCarthy et al. 2007), but population genetic analysis has revealed that the majority of human cases of disease are caused by isolates associated with livestock and poultry (Wilson et al. 2008).

Qualitative evidence suggests that adaptation is ongoing in *C. jejuni* and is facilitated by horizontal gene transfer. The mechanism of virulence in *C. jejuni* is still poorly understood, but loci responsible for adherence, cellular invasion, toxin production, and flagellar motility are thought to be important virulence factors (Fouts et al. 2005). The genetic basis for antimicrobial drug resistance is known, and its spread by recombination has been demonstrated both within *C. jejuni* (de Boer et al. 2002) and between related species (Oyarzabal et al. 2007). The resistance of *C. jejuni* to a range of antibiotics is common throughout the world and is thought to have been driven by frequent use in animals farmed for meat (Moore et al. 2006).

What we do not know about *C. jejuni* is how dynamic is it as a species, what is the timescale of its evolution, how quickly might it adapt, and what is the extent to which recombination facilitates gene transfer within *C. jejuni* and to or from its sister species? Recently, Sheppard et al. (2008) have suggested that the rate of recombination between *C. jejuni* and its sister species *Campylobacter coli* is sufficient to have begun to reverse the speciation process, but the timescale over which that might be happening is unclear.

Integral to these questions is the matter of calibrating the molecular clock, an issue fraught with difficulty in the bacterial world. Unlike multicellular eukaryotes, bacteria do not easily fossilize, and when they do their unremarkable morphology does not allow accurate taxonomic classification. Unlike viruses, bacteria mutate slowly, so slowly that the evolution of natural populations has not been readily measured in real time. To date, the rate of evolution in bacteria has been calibrated indirectly. Ochman and Wilson (1987) placed upper and lower bounds on a series of bacterial phylogenetic splits by cross-referencing other events that can be dated. For example, the common ancestor of mitochondria and their closest living bacterial relatives must have occurred more recently than the Palaeoproterozoic increase in atmospheric oxygen and prior to the radiation of mitochondrion-bearing eukaryotes. Moran et al. (1993) calibrated the molecular clock in bacterial endosymbionts by assuming

[1] Present address: Department of Human Genetics, University of Chicago, 920 East 58th Street, CLSC 410, Chicago, IL 60637 USA.
[2] Present address: Université d'Avignon, IUT STID, Site Agroparc, BP 1207, Avignon, France.
[3] Deceased.

Key words: *Campylobacter jejuni*, molecular clock, recombination, selection, coalescent, Neolithic.

E-mail: djw@uchicago.edu.

Mol. Biol. Evol. 26(2):385–397. 2009
doi:10.1093/molbev/msn264
Advance Access publication November 13, 2008

cospeciation with their aphid hosts, for whom a fossil record is available. These phylogenetic approaches often conflict with empirical approaches that are based on laboratory measurements of generation lengths and mutation rates (Lenski et al. 2003; Ochman 2003).

Our study is based on a longitudinal sample of 1,205 *C. jejuni* isolates collected over a 3-year period from patients in Lancashire, England (Wilson et al. 2008), and DNA sequenced using multilocus sequence typing (MLST, Dingle et al. 2001). We characterize the ongoing evolution of *C. jejuni* using a population genetic (microevolutionary) model of the forces of drift, mutation, recombination, and natural selection. We exploit the longitudinal sample to calibrate the molecular clock for *C. jejuni* directly from within-species variation, and we utilize empirical measurements of generation lengths and mutation rates to quantify the species' effective population size and evolutionary potential. Finally, we look at the evolution of *C. jejuni* in the wider (macroevolutionary) context of the *Campylobacter* genus, employing our rate estimates to date phylogenetic splits, such as the common ancestor of *C. jejuni* and its sister species *C. coli*. We evaluate our rate estimates in light of previous work and discuss the plausibility of a Neolithic origin of *C. jejuni*.

Methods

Isolates

We analyzed 1,205 of the *C. jejuni* isolates of Wilson et al. (2008) that were available at the time of writing. The isolates were collected from patients diagnosed with campylobacteriosis and notified through general practitioners and hospitals to the Preston Microbiology Services Laboratory in the Preston postcode district between January 1, 2000, and December 31, 2002. The sampling time corresponds to the date at which the isolate was received at the laboratory. The study area covers 968 km^2, consisting of both urban (Preston, Leyland, Chorley, and Garstang) and rural (Ribble estuary and Ribble valley) districts and comprised 403,000 people at the 2001 census. As is the norm with campylobacteriosis, the cases we studied were sporadic in nature; there was no evidence for outbreaks. The isolates were sequenced at seven housekeeping loci (*aspA*, *glnA*, *gltA*, *glyA*, *pgm*, *tkt*, and *uncA*) using the MLST (Dingle et al. 2001) producing 3,309 nucleotides in total per isolate.

Analysis Overview

We used a variety of standard tools from a molecular evolution toolkit in order to infer the evolutionary history of *C. jejuni*: Structure (Falush et al. 2003) to identify *C. jejuni–C. coli* hybrids, an importance sampler (Fearnhead 2008) to calibrate the rate of molecular evolution, approximate Bayesian computation (ABC) (Beaumont et al. 2002) to estimate population genetic parameters for *C. jejuni*, and BEAST (Drummond et al. 2002) to reconstruct the phylogeny of the genus *Campylobacter*.

Identification of *C. coli* Hybrids

It was necessary to identify *C. jejuni–C. coli* hybrids because the microevolutionary models that we subsequently used are based on the coalescent (Kingman 1982), which is a probabilistic model for the relatedness of individuals in a population (the genealogy). Interspecific gene flow introduces alleles that are very distantly related to the rest of the population, and therefore violate the assumptions of relatedness made by the coalescent. Gene flow between *C. jejuni* and its sister species *C. coli* has been reported previously (Dingle et al. 2005), so we used Structure (Falush et al. 2003) to identify alleles imported from *C. coli* before fitting the evolutionary models.

Informally, Structure assigns individuals to populations (or species in our case) on the basis of allele frequencies. Alleles at one locus that are typically observed to be associated with *C. coli* alleles at other loci will be assigned to *C. coli*. Although they are sister species, relatively few alleles are found in both species. Except when there has been interspecific gene flow, there are usually enough fixed nucleotide differences between *C. jejuni* and *C. coli* alleles to accurately identify the origin (Dingle et al. 2001). For each polymorphic nucleotide, Structure gives a posterior probability of *C. jejuni* ancestry (as opposed to *C. coli* ancestry). From this we defined hybrids as isolates with a posterior probability of dual ancestry greater than 0.95. Full details of the analysis are given in the supplementary methods, Supplementary Material online.

Microevolution of *C. jejuni*

The evolutionary models that we employed, both population genetic and phylogenetic, are modular insofar as they comprise the following components:

- Model of mutation
- Model of recombination
- Model of relatedness (the genealogy or phylogeny)
- Method of statistical inference

In evolutionary genetics, the choice of model represents a tradeoff between the competing desires for a biologically realistic model and one for which statistical inference is feasible. The method of inference is the limiting step: Simpler models can be fitted using powerful likelihood-based methods (such as importance sampling [IS] or Markov chain Monte Carlo [MCMC]), which are statistically efficient in that they exploit all the information the data have to offer. More complex models can only be fitted using simulation-based methods (such as ABC, Beaumont et al. 2002), which are suboptimal because they use summaries of the full data.

Recombination in particular massively increases the complexity of a model, and intraspecific recombination is frequent in *C. jejuni* (Fearnhead et al. 2005), both within and between genes. To learn about the evolution of *C. jejuni*, we adopted a two-stage approach that represents a compromise between our competing desires 1) to calibrate the rate of evolutionary change in *C. jejuni* and 2) to learn about the multifarious forces shaping *C. jejuni*. The former requires powerful likelihood-based inference to extract what is likely to be a weak signal, because our sampling period of 3 years is likely to be short relative to the timescale of bacterial evolution. The latter requires a complex model that incorporates drift, mutation, selection, and recombination.

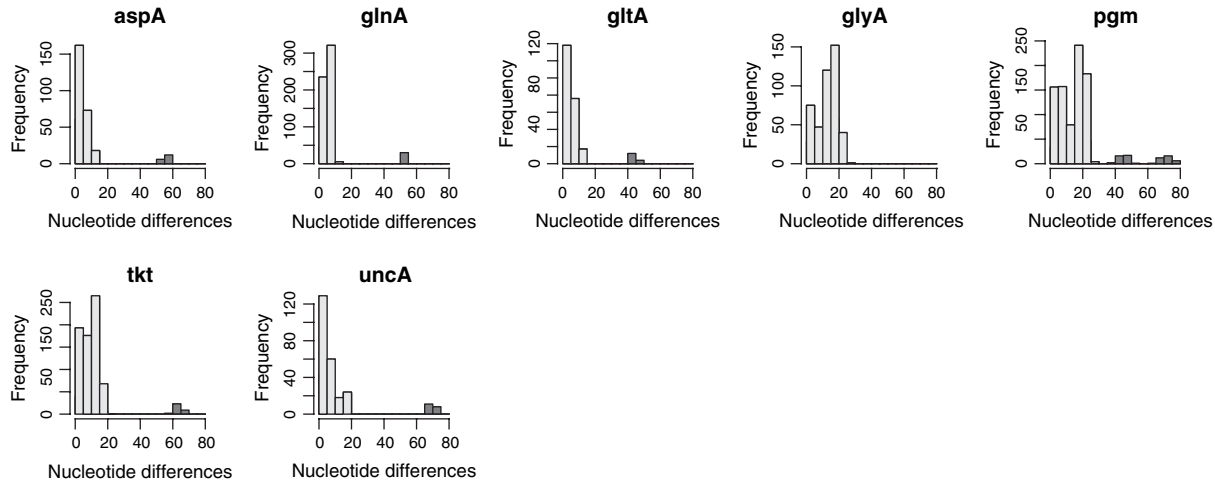

FIG. 1.—Histograms of the number of nucleotide differences between each pair of alleles at the seven MLST loci. When *Campylobacter coli*–derived alleles are removed, the dark gray portions of the histograms disappear.

In the first step, we fitted the following model that is simple enough to use a likelihood-based IS method (Example 3 of Fearnhead 2008). The 1,018 isolates were analyzed for which sampling times were available.

- Infinite alleles model of mutation (Kimura and Crow 1964), in which new alleles are generated at rate θ. This rate encompasses the generation of novel alleles by mutation, intraspecific recombination, and interspecific recombination.
- Free recombination between loci, which is to say the loci are assumed unlinked or independent.
- The coalescent with serial samples (Rodrigo and Felsenstein 1999), in which the parameter $N_e g$ determines the rate of coalescence per year.
- Likelihood-based IS (Fearnhead 2008).

The allelic mutation rate θ is measured in coalescent time units of $2N_e g$ years, where N_e is the effective population size and g the generation length in years. In this model, the alleles identified as *C. coli* imports were assumed to have arisen by interspecific recombination, and the dates of their introduction were estimated. We used priors on θ and $N_e g$ that are flat on the logarithmic scale. The object of inference in this, the first step, was to estimate the timescale of evolution, which is determined by the coalescent parameter $N_e g$. For an illustration of how $N_e g$ affects the signal of measurable evolution, see supplementary figure S1, Supplementary Material online.

In the second step, we fitted a more complex model using ABC (see supplementary methods, Supplementary Material online, for full details), in which selection is modeled as a form of mutational bias. This analysis was based on the sequences of 881 pure *C. jejuni* isolates available at the time.

- Nielsen and Yang (1998) codon model whose parameters are the synonymous mutation rate θ_S, the transition–transversion ratio κ, and the d_N/d_S ratio ω.
- A model of recombination suitable for bacteria (Wiuf and Hein 2000), whose parameters are the rate of recombination ρ and the average length of import τ.
- The coalescent with serial samples, with parameter $N_e g$.
- Simulation-based ABC.

The parameters θ_S and ρ are measured per kilobase (kb), in coalescent time units of $2N_e g$ years. In this model, *C. coli* hybrids were excluded from the analysis. Except for $N_e g$, we used priors that are flat on the logarithmic scale. In order to calibrate the rate of molecular change, which is to say convert θ_S into real-time units of years, we used the posterior from the simpler model as an informative prior for $N_e g$. This was necessary because we had found that ABC was not sufficiently sensitive to estimate $N_e g$.

To distinguish the calibrated parameters from those measured in coalescent time units, we use μ_S to denote the rate of synonymous mutation per kb per year and r to denote the rate of recombination per kb per year, whereas θ_S and ρ denote the corresponding rates in coalescent time units. Formally, $\theta_S = 2N_e g \mu_S$ and $\rho = 2N_e g r$. We also use θ_N and μ_N to denote the corresponding rates of nonsynonymous change, and the total mutation rates are $\theta = \theta_S + \theta_N$ and $\mu = \mu_S + \mu_N$.

Macroevolution of *Campylobacter*

We put the evolution of *C. jejuni* into a wider context by inferring the phylogenetic history of seven *Campylobacter* species for which similar MLST schemes have been designed (Dingle et al. 2001; Miller et al. 2005; van Bergen et al. 2005). For each species we chose a typical isolate and tested that there was no interspecies recombination between the chosen isolates using a permutation test based on the correlation between physical distance and linkage disequilibrium (LD) (McVean et al. 2002). We then fitted the following phylogenetic model to the concatenated gene sequences. For further details see supplementary methods, Supplementary Material online.

- Nielsen and Yang (1998) codon model with parameters μ_S, κ, and ω.
- No recombination, which is to say the loci are fully linked.

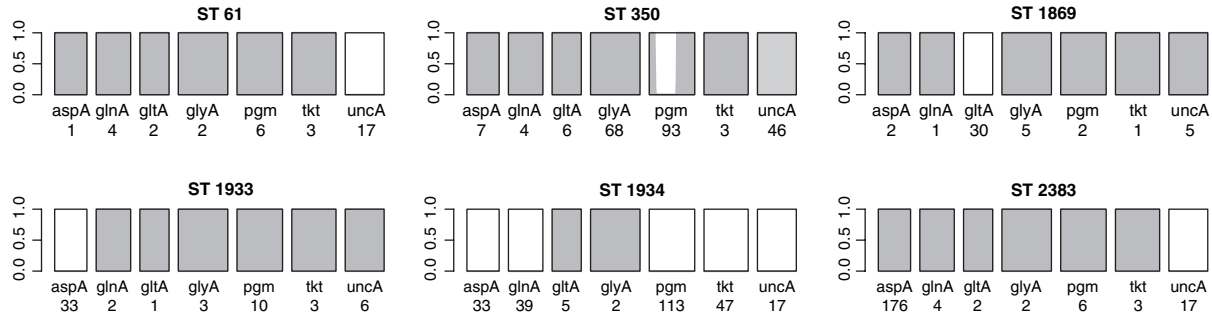

FIG. 2.—*Campylobacter jejuni–Campylobacter coli* hybrids identified by Structure. Six sequence types were identified as hybrids. For each hybrid, the posterior probability of *C. jejuni* ancestry (as opposed to *C. coli* ancestry) is shown in gray for the seven MLST loci.

- Yule (1924) model, in which speciation occurs at rate λ per 1,000 years.
- Likelihood-based MCMC implemented in BEAST (Drummond et al. 2002).

We used informative priors for κ and ω, taken from our posteriors estimated from the microevolutionary model. On the timescale of the *Campylobacter* phylogeny, our sequences are essentially sampled contemporaneously. Therefore, the data do not contain information regarding the rate of evolution. That was provided entirely by our informative prior on μ_S, which was taken from the posterior estimated from the microevolutionary model.

Results

In presenting our results, we begin by scrutinizing genetic diversity and LD in the contemporary population. This reveals a number of insights, including the identification of new *C. jejuni–C. coli* hybrids, and the demonstration that recombination is the primary mechanism driving molecular change. Then we proceed to calibrating the real-time rate of molecular change. We detect a signal of measurable evolution in *C. jejuni* and employ that calibration to date historical events, including the importation of genes from *C. coli* and the most recent common ancestor (MRCA) of *C. jejuni*. We discuss the real-time evolutionary potential of *C. jejuni*, including the likely efficacy of selection and the size of the gene pool. Finally, we utilize our estimate of the molecular clock to calibrate the phylogeny of the genus *Campylobacter*. We discuss the cultural changes that may have coincided with the *C. jejuni–C. coli* split and the robustness of the approach in light of the conflict that arises with traditional estimates.

Recombination Dominates the Evolution of *C. jejuni*

A cursory analysis of the patterns of nucleotide diversity in *C. jejuni* immediately reveals evidence of interspecies gene flow. For each of the seven genes, figure 1 illustrates the number of nucleotide differences between each pair of alleles. At six loci, there is distinct clustering between alleles that are genetically similar (in light gray) and dissimilar (in dark gray). The majority of alleles differ by fewer than 20 nucleotides, but a small number are highly divergent, differing by 40 nucleotides or more from the rest. By comparing our sample of *C. jejuni* isolates to the *C. coli* isolates of Dingle et al. (2005) using the program Structure (Falush et al. 2003), we found that the genetically divergent alleles, most of which are observed at low frequency, are imports from *C. coli*.

Thirty of 1,205 *C. jejuni* isolates were found to contain *C. coli* alleles. Sequence type (ST-) 61, which carries the *C. coli*-derived *uncA*-17 allele (Dingle et al. 2005), accounted for 23 of those. Five other STs were newly identified as hybrids; for each one, figure 2 illustrates the posterior probability of *C. jejuni* ancestry across the seven loci. ST-1869 and ST-1933, represented by one isolate each, carried *C. coli*–derived alleles *gltA*-30 and *aspA*-33, respectively. ST-2383, also represented by a single isolate, differed from ST-61 by a single nucleotide in *aspA*. The only evidence for intragenic mosaicism was found in *pgm*-93, carried by ST-350, of which there were two isolates. This allele contains a 200-bp stretch of DNA with *C. coli* ancestry, flanked on both sides by sequence of *C. jejuni* ancestry. The rarity *aspA*-33, *gltA*-30, and *pgm*-93, and the uniformity of the background on which *uncA*-17 is found, suggest these genes were introduced singly in four independent ancestral cross-species recombination events. However, the ancestral history of ST-1934, represented by a single isolate, is likely to be more complex. Five of the seven genes were *C. coli* derived, intimating a more complex history involving multiple cross-species gene transfer events. A single isolate, ST-962, bore *C. coli* alleles at all seven genes, suggesting this isolate was misclassified.

Removal of the *C. coli* alleles erases the genetic discontinuity from patterns of *C. jejuni* diversity (leaving the light gray portions of fig. 1) and allows us to identify the contribution of mutation and selection to population diversity. Pure *C. jejuni* isolates differed by an average of 49.8 nucleotides out of the 3,309 sequenced by MLST. The majority of differences were synonymous, indicative of the housekeeping function of the loci. Using ABC, we estimated that a pair of *C. jejuni* isolates, sampled at the same time, differ by $\theta = 13.7$ mutations per kb on average, of which $\theta_S = 11.8$ would be synonymous (see table 1 for credible intervals [CI]). This makes *C. jejuni* diverse relative to other bacteria (Pérez-Losada et al. 2006). The intense purifying selection experienced by these genes is reflected in the small d_N/d_S ratio of $\omega = 0.0283$. Mutations are also highly skewed in favor of transitions ($\kappa = 19.0$). To emphasize the strength of selection, we used our parameter

Table 1
Estimates of Evolutionary Parameters in *Campylobacter Jejuni*

Parameter		Units	A Posteriori		A priori 95% CI
			Point Estimate	95% CI	
Allele model					
$N_e g$	Mean coalescence time	Years	209	155–288	92–820
θ	Allelic mutation rate	$(2N_e g)^{-1}$	4.38	3.75–5.18	1.1–9.4
1					
Sequence model					
θ	Total mutation rate	kb^{-1} $(2N_e g)^{-1}$	13.7	8.35–23.9	2.1–180
θ_S	Synonymous mutation rate	kb^{-1} $(2N_e g)^{-1}$	11.8	7.51–19.7	1.9–170
θ_N	Nonsynonymous mutation rate	kb^{-1} $(2N_e g)^{-1}$	1.9	0.84–4.2	0.2–10
θ_0	Neutral mutation rate	kb^{-1} $(2N_e g)^{-1}$	35.6	22.0–61.8	5.8–510
κ	Transition–transversion ratio		19.0	8.85–39.8	3.3–180
ω	d_N/d_S ratio		0.0283	0.0165–0.0492	0.0022–0.18
$\rho\tau$	Recombination rate between distant loci	$(2N_e g)^{-1}$	6.08	3.19–11.2	2×10^{-4}–8×10^{5}
ρ	Recombination rate	kb^{-1} $(2N_e g)^{-1}$	1.31	0.0273–42.2	0.0014–72
τ	Mean DNA import length	kb	4.54	0.100–214	0.015–6,800

The geometric mean was used to obtain point estimates. The (2.5%, 97.5%) quantiles were used to calculate the 95% CI. All priors were uniform on the logarithmic scale.

estimates to calculate that the theoretical mutation rate, in the absence of selection (i.e., if $\omega = 1$), would be $\theta_0 = 35.6$ per kb, some 2.6 times higher than the actual rate. This implies that purifying selection purges 60% of novel genetic variation.

Recombination is prevalent in *C. jejuni* (Fearnhead et al. 2005), facilitating gene flow within the species, as well as importing diversity from without. Permutation tests (McVean et al. 2002) showed significantly lower LD between loci than within $P < 0.01$), but lacked power to demonstrate significant levels of intragenic recombination on a locus-by-locus basis. Nevertheless, formal model fitting revealed nonzero levels of intragenic recombination. Using ABC, we estimated that an average of $\rho = 1.31$ recombination breakpoints per kb would occur on the evolutionary branches separating a pair of isolates. Assuming an exponential distribution for the length of DNA imported during homologous recombination, we estimated an average import length of $\tau = 4.54$ kb. The wide CIs for ρ and τ (table 1) attest to the weak intragenic signal, but these figures are consistent with those estimated by LDhat (Fearnhead et al. 2005). Appreciably more power was available to estimate the interlocus rate of recombination. A curious aspect of bacterial recombination is that residual, nonzero LD is expected even between distant loci (Wiuf and Hein 2000). This is because the rate of recombination plateaus to a maximum of $\rho\tau$, rather than increasing linearly with physical distance. We estimated the long-range rate of recombination to be $\rho\tau = 6.08$, with a relatively tight CI (see table 1).

A simple comparison of the rate of recombination to mutation conceals the primacy of recombination as the dominant force that generates molecular change. The higher rate of point mutation per se is offset by the wider effect that each recombination event has. Hundreds to thousands of nucleotides are imported during recombination, of which a proportion π/1,000 will differ between the incoming and existing sequence, where π is the average number of pairwise nucleotide differences per kb in the population. Therefore, within-species recombination drives molecular change by a factor $\rho\tau\pi/2\theta$ more quickly than de novo mutation, which we estimated to equal 2.67 (95% CI 1.39–4.95).

By quantifying the rate of recombination, we were able not only to establish the importance of intraspecific recombination relative to mutation but also to compare the relative importance of intraspecific with interspecific gene flow. We reasoned earlier that the ancestral history of *C. jejuni* bears witness to at least four importations from sister species *C. coli*. From our estimate of the within-species recombination rate, we predict that during the same ancestral history, there were approximately 230 intraspecific recombination events. Although we did not formally infer the rate of recombination between *C. jejuni* and *C. coli*, we can deduce that the rate of cross-species gene flow is little more than an order of magnitude (roughly 230/4 = 57.5 times) less frequent than within-species gene flow. This reinforces the observation that recombination is fundamental, not just in driving molecular change within *C. jejuni* but also in facilitating cross-species gene flow, which is likely to have important implications for long-term adaptation.

Is Genetic Variation in *C. jejuni* Just 400 Years Old?

Central to calibrating the rate of molecular evolution is estimating $N_e g$, the timescale of the genealogy. The product of the effective population size (N_e) and the generation length (g), this parameter dictates the rate of coalescence in *C. jejuni*, it determines the date of the MRCA, and it allows the rates of mutation and recombination to be measured in real-time units. To have power to estimate $N_e g$, the population must be measurably evolving on the timescale of the sampling period (Drummond et al. 2003). That is equivalent to saying there must have been a detectable number of mutation, recombination, or coalescence events in the population during the 3-year longitudinal study. Supplementary figure S1, Supplementary Material online, illustrates the idea and shows where in the data the signal lies.

We used the importance sampler of Fearnhead (2008) to estimate the rate of molecular change in *C. jejuni* (table 1). To determine whether the population was measurably evolving, we conducted a formal hypothesis test in which we compared two models, one using the longitudinal sampling times and the other assuming all sequences were

Table 2
Dating Ancestral Events

Event		Calibrated Date		Uncalibrated Date (Units of $N_e g$)	
		Point Estimate	95% CI	Point Estimate	95% CI
T_{MRCA}	MRCA	1,592	1,305–1,772	1.87	9.21×10^{-1}–3.68
$T_{aspA\text{-}33}$	Import of *aspA-33* from *Campylobacter coli*	August 2000	Jan 1997–Mar 2001	1.34×10^{-3}	8.68×10^{-5}–0.0422
$T_{pgm\text{-}93}$	Import of *pgm-93* from *C. coli*	January 1998	Jul 1965–May 2000	8.73×10^{-3}	8.25×10^{-4}–0.244
$T_{uncA\text{-}17}$	Import of *uncA-17* from *C. coli*	March 1966	Sep 1726–Jul 1996	1.20×10^{-1}	1.23×10^{-2}–1.06

sampled simultaneously. For the two models, measurably evolving (M_1) versus not measurably evolving (M_0), we estimated the likelihood averaged over the parameter values and calculated a Bayes factor, which is the ratio of these likelihoods. We obtained a Bayes factor of 3.0×10^{20}, which is much greater than one, indicating very strong support for the measurably evolving hypothesis.

We estimated that $N_e g$ 209 years (95% CI 155–288), which can be understood as the average age of the common ancestor of a pair of *C. jejuni* sequences sampled contemporaneously. The magnitude of $N_e g$, together with the fact that the population is measurably evolving on a timescale of just 3 years, suggests a rapid rate of evolution in *C. jejuni*. In a demographically stable population, the date of the MRCA is a measure of population turnover and is expected to equal $2N_e g$. In a recombining population such as *C. jejuni*, the MRCA varies between genes and even within a gene. We estimated a very recent date for the average MRCA across genes of T_{MRCA} = AD 1591 (95% CI 1304–1771; see table 2). This suggests that the rate of turnover of genetic variation at the average locus is just 409 years. The MRCA represents merely the root of the contemporary genealogy and does not coincide with the birth of the species, but it does mark a horizon beyond which we cannot use intraspecific genetic variation to reconstruct the evolutionary past. Later, we employ homologous sequences from other *Campylobacter* species to infer the deeper history of *C. jejuni*.

Evaluating the timescale of evolution in *C. jejuni* should allow events other than the MRCA to be dated. Of particular interest are the dates of the cross-species recombination events identified earlier. Those events are, namely, the importation of *aspA*-33, *gltA*-30, *pgm*-93, and *uncA*-17. We report both calibrated and uncalibrated dates in table 2; the uncalibrated dates are measured in coalescent time units of $N_e g$. We found that two genes are very recent imports from *C. coli*, consistent with their low frequency. The calibrated dates suggest that with 95% probability, *aspA*-33 was introduced between January 1997 and March 2001, and *pgm*-93 was imported between July 1965 and May 2000. The uncalibrated dates show that these rare alleles were imported very much more recently than the T_{MRCA}. We were unable to date the importation of *gltA*-30 because no sampling time was recorded for the isolate carrying it. However, based on its sample frequency, its importation date is likely to resemble that of *aspA*-33. The *C. coli*-derived *uncA*-17 is the most frequent and most ancient import. Present in 25 isolates, we estimated its date of introduction to be March 1966 (95% CI September 1725–July 1995) or 6.5% of the evolutionary time since the MRCA.

Knowledge regarding the timescale of evolution in *C. jejuni* allows the absolute rate of molecular change, or the molecular clock, to be calibrated. The parameter θ is the number of mutations (per kb) by which the average pair of sequences differs. However, an equivalent interpretation is that θ is the rate of mutation (per kb) in coalescent time units of $2N_e g$ years. Similarly, ρ is the rate of recombination (per kb) per $2N_e g$ years. Therefore, we can combine our estimate of $N_e g$ obtained by IS with our estimates of population genetic parameters obtained by ABC to convert the mutation and recombination rates from coalescent time to real time. Table 3 summarizes the results of this conversion. We inferred an absolute mutation rate of $\mu = 3.23 \times 10^{-2}$ per kb per year, comprising a synonymous rate of $\mu_S = 2.79 \times 10^{-2}$ and a nonsynonymous rate of $\mu_N = 4.4 \times 10^{-3}$ per kb per year. This corresponds to an average waiting time, per lineage per kb, of 31.0 years before a mutation arises. We estimated an absolute recombination rate of $r = 3.07 \times 10^{-3}$ per kb per year, rising to $r\tau = 1.45 \times 10^{-2}$ per year between distant loci. We go on to use the absolute rate of synonymous mutation (μ_S) calculated here to calibrate the phylogeny of the genus *Campylobacter*.

We have already seen that *C. jejuni* experiences intense purifying selection at the housekeeping loci studied here ($\omega = 0.0273$). In the absence of selection, the absolute mutation rate would be $\mu_0 = 8.39 \times 10^{-2}$ per kb per year (95% CI $4.85 \times 0^{-2} - 15.1 \times 10^{-2}$), 2.6 times higher than that observed. Besides reflecting the strict functional constraint of the genes in question, the strength of selection conveyed by these figures intimates a large effective

Table 3
Calibrated Rate Parameters in *Campylobacter jejuni*

Parameter		Units	Point Estimate	95% CI
μ	Total mutation rate	kb^{-1} $year^{-1}$	3.23×10^{-2}	1.86×10^{-2}–5.81×10^{-2}
μ_S	Synonymous mutation rate	kb^{-1} $year^{-1}$	2.79×10^{-2}	1.60×10^{-2}–5.08×10^{-2}
μ_N	Nonsynonymous mutation rate	kb^{-1} $year^{-1}$	4.40×10^{-3}	2.60×10^{-3}–7.30×10^{-3}
μ_0	Neutral mutation rate	kb^{-1} $year^{-1}$	8.39×10^{-2}	4.85×10^{-2}–1.51×10^{-1}
$r\tau$	Recombination rate between distant loci	$Year^{-1}$	1.45×10^{-2}	7.11×10^{-3}–2.92×10^{-2}
r	Recombination rate	kb^{-1} $year^{-1}$	3.07×10^{-3}	4.79×10^{-5}–1.27×10^{-1}

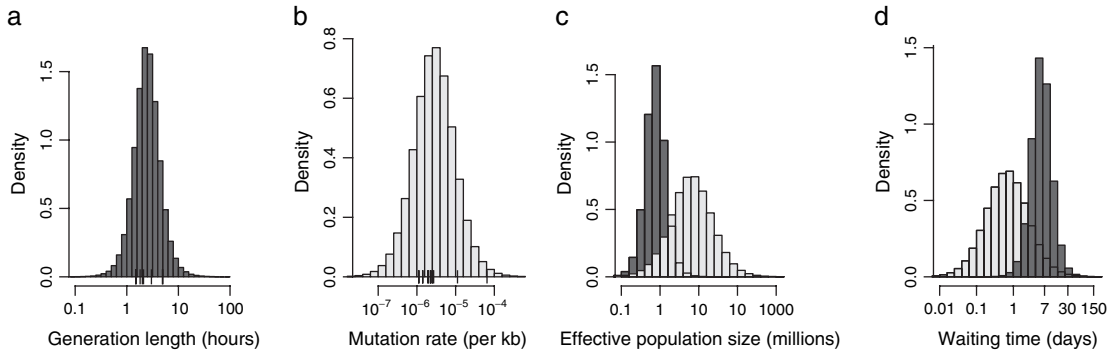

FIG. 3.—Inference based on empirical parameter estimation. Posterior distributions of (*a*) generation length g, (*b*) mutation rate m_0, (*c*) effective population size N_e, and (*d*) waiting time for a novel mutation W, based on empirical estimates of the generation length (dark gray histograms) or mutation rate (light gray histograms), and population genetic estimate of θ_0 (*c*, *d*). The empirical data are indicated with black lines crossing the horizontal axis in (*a*) and (*b*). The background histogram in (*c*) and (*d*) additionally depends on the population genetic estimate of $N_e g$, rendering them sensitive to the calibration of the molecular clock. See table 4 for more details.

population size and gene pool. Intense functional constraint suggests that *C. jejuni* is already highly adapted, but it is of interest to quantify the evolutionary potential of the species or the rate at which it could respond to a change in selection pressure.

The Evolutionary Potential of *C. jejuni* Is Immense

The evolutionary potential, or adaptability, of a species depends on two quantities not readily calculable from population genetic data. The real-time total-population rate of mutation, $N_e\mu_0$, determines the rate at which advantageous variants may arise in the event of a change in selection pressure. Together with the amount of standing variation, it determines the rate of adaptation; informally, this is the size of the gene pool. The effective population size, N_e, limits the efficacy of selection (Kimura 1955) and hence how likely it is that an advantageous variant would spread should it arise. We cannot estimate either quantity directly from the sequence data. However, knowledge of the generation length (g) or the per-generation neutral mutation rate (m_0) would be sufficient to calculate the quantities of interest, via the parameters we have been able to infer.

Campylobacter jejuni is a microaerophilic bacterium that is adapted to growth at 37 or 42 °C, typical of the mammalian and avian gut, respectively. Growth experiments in culture demonstrate that the generation length depends on many factors including the genotype, temperature, and the presence of competing strains (Velayudhan and Kelly 2002; Khanna et al. 2006; Jackson et al. 2007; Konkel et al. 2007). An excursion into the recent literature reveals that the doubling time of wild type *C. jejuni* ranges from 90 min to 5 h, although we cannot discount longer generation times in vivo. By assuming that empirical estimates of generation length follow a log-normal distribution and utilizing vague priors on the mean and variance of that distribution, we obtained a posterior-predictive distribution on generation length (see, e.g., Gelman et al. 2003, p. 74). Figure 3*a* shows this distribution, with the empirically measured generation times from 10 growth experiments (Velayudhan and Kelly 2002; Khanna et al. 2006; Jackson et al. 2007; Konkel et al. 2007) of wild type *C. jejuni* at 37 and 42 °C, both alone and in mixed culture, indicated by black lines crossing the horizontal axis. We calculated a point estimate of 2.44 h, with 95% CI of 0.719–8.29 h (table 4).

No direct estimate of the de novo per-generation mutation rate, m_0, is available for *C. jejuni*, but Drake (1991) and Drake et al. (1998) reported a startlingly consistent pattern in the genomic rate of mutation between different micro-organisms. Per kb, the mutation rates of these organisms (bacteriophages, *E. coli*, *Saccharomyces cerevisiae*, and *Neurospora crassa*) vary 16,000-fold. Per genome, the rate varies just 2.5-fold, excepting outliers. Drake calculated a mean genomic mutation rate of

Table 4
Inference Based on Empirical Parameter Estimates

Parameter		Units	Method	Point Estimate	95% CI
Empirical estimates					
g	Generation length	Years	In vitro doubling times	2.79×10^{-4}	8.20×10^{-5}–9.46×10^{-4}
		Hours		2.44	0.719–8.29
m_0	Neutral mutation rate	$kb^{-1}\ g^{-1}$	Drake's method	2.77×10^{-6}	2.26×10^{-7}–3.41×10^{-5}
Composite estimates					
N_e	Effective population size		$N_e = \theta_0/(2m_0)$	6.42×10^6	4.95×10^5–8.29×10^7
			$N_e = N_e g/g$ †	7.50×10^5	2.13×10^5–2.64×10^6
W	Inverse population mutation rate	Days	$W = 365 \times (2g)/\theta_0$	5.71	1.52–21.4
		Days	$W = 365 \times (4N_e g m_0)/\theta_0{}^2$ †	0.667	0.0437–10.2
μ_0	Neutral mutation rate	$kb^{-1}\ year^{-1}$	$\mu_0 = m_0/g$	9.95×10^{-3}	6.18×10^{-4}–1.60×10^{-1}

Methods denoted by † are calibration sensitive.

3.3×10^{-3}, which equals 1.9×10^{-6} per kb for *C. jejuni*, which has a 1.7-Mb genome (www.nmpdr.org/content/campy.php). To quantify the uncertainty in this approach, we again assumed a log-normal distribution for variation in m_0 to obtain a posterior-predictive distribution based on Drake's mutation rates, adjusted for genome size. Because we included outliers, we obtained a point estimate of 2.77×10^{-6} per kb and wide CI (table 4). The distribution is plotted in figure 3*b*, with the empirically measured mutation rates, adjusted for genome size, indicated by black lines crossing the horizontal axis.

From these empirical estimates, we can calculate evolutionary quantities of interest. A number of simple formulae relate the parameters g, m_0, θ_0, and $N_e g$ to the effective population size, N_e, and the real-time population mutation rate, $N_e\mu_0$. An intuitive representation of this latter quantity is $W = 365 \times 1{,}000/(N_e\mu_0)$, which gives the expected waiting time, in days, for a mutation to arise at any particular nucleotide, somewhere in the population. The formulae for calculating N_e and W are detailed in table 4. When the formula involves $N_e g$, the estimate is sensitive to the calibration of the molecular clock: N_e is calibration sensitive when estimated from g but not m_0. W is calibration sensitive when estimated from m_0 but not g. By comparing calibration-sensitive and insensitive estimates, this provides a useful check.

The posterior distribution of N_e is plotted as a light gray histogram in the foreground of figure 3*c* to emphasize its dependence on m_0 and its insensitivity to calibration. In the background, in dark gray is plotted the estimate based on g, which is calibration sensitive. Based on m_0, we estimate an effective population size of 6.42 million. The wide CIs (table 4) principally reflect the underlying uncertainty in m_0. The calibration-sensitive estimate based on g is 0.75 million and has tighter, partially overlapping CIs that reflect the lesser uncertainty in g. In figure 3*d*, the estimate of W based on g (dark gray histogram in foreground) is insensitive to calibration and has tight CIs surrounding the point estimate of 5.71 days. The estimate based on m_0 is 0.667 days (light gray histogram in background); the posterior distribution has wide CIs and is calibration sensitive.

Qualitatively, the results agree that the effective population size of *C. jejuni* is large, on the order of hundreds of thousands to tens of millions. This suggests that selection is highly efficacious in *C. jejuni*, sensitive to fitness advantages as small as 1×10^{-6}. Earlier, we estimated a d_N/d_S ratio of 0.0283, suggesting that selection is, for the most part, purifying. However, a second consequence of the large effective population size is to counteract the low intrinsic mutation rate, such that the average waiting time for a new mutation to arise at any particular nucleotide is just 5.7 days, maybe less. Although most are lost by drift or purged by selection, novel genetic variants are arising at such a rate that the potential of *C. jejuni* to adapt to changes in selection pressure appears to be immense.

Despite the qualitatively similar conclusions, there exist inconsistencies between the estimates of N_e and W based on g and m_0. These inconsistencies are manifest in the nonoverlapping portions of the light and dark gray histograms in figure 3*c* and *d*. The reason for this partial overlap can be understood when we use g and m_0 to obtain an empirical calibration of the neutral rate of molecular change in *C. jejuni*. Table 4 shows that the empirical point estimate of μ_0 is 9.95×10^{-3}, an order of magnitude lower than that estimated by our population genetics method from within-species variation (table 3). Although the CIs of the empirical estimate are sufficiently wide (table 4) to subsume the population genetics estimate, the empirical evidence suggests a slower molecular clock overall.

Is *Campylobacter* Speciating on a Timescale of Thousands of Years?

Earlier, we remarked that the MRCA of a species constitutes a horizon beyond which we cannot use intraspecific genetic variation to reconstruct evolutionary history. In a recombining species, the MRCA can differ between loci, but we estimated that the date of the MRCA for the average locus in *C. jejuni* existed around 400 years ago. To delve deeper into the species' evolutionary history, it is necessary to employ more distantly related molecular sequences, so we used the sequences of six *Campylobacter* species for which MLST schemes have also been developed: *C. coli* (Dingle et al. 2001), *Campylobacter fetus* (van Bergen et al. 2005), *Campylobacter helveticus* (Miller et al. 2005), *Campylobacter insulaenigrae* (Stoddard et al. 2007), *Campylobacter lari* (Miller et al. 2005), and *C. upsaliensis* (Miller et al. 2005). The MLST schemes for these species have four genes in common: *glnA*, *glyA*, *tkt*, and *uncA* (otherwise known as *atpA*).

The *Campylobacter* species we studied exhibit an interesting array of pathogenicities and host ranges. Besides *C. jejuni*, which is responsible for 90% of human campylobacteriosis (Gillespie et al. 2002), *C. coli*, *C. lari*, and *C. upsaliensis* have been documented in sporadic cases or outbreaks of gastroenteritis in humans (Miller et al. 2005). *Campylobacter coli* has a host range largely overlapping with that of *C. jejuni* albeit with a greater affinity for pigs (Dingle et al. 2005) and, like its sister species, tends to be carried asymptomatically in the gut of livestock, poultry, wild birds, and mammals. *Campylobacter lari* likewise has a wide host range but is characterized by its isolation from seagulls, mussels, and oysters (Miller et al. 2005). *Campylobacter upsaliensis*, together with the related *C. helveticus*, is associated with domestic cats and dogs (Miller et al. 2005). Two subspecies are known of *C. fetus*. *C. fetus* subsp. *fetus* can induce abortion in sheep and less often in cattle and humans (van Bergen et al. 2005). The genetically uniform *C. fetus* subsp. *venerealis* is cattle restricted, in which it causes a venereal infection that can lead to infertility and abortion (van Bergen et al. 2005). A somewhat distinct, reptile-associated strain of *C. fetus* has also been described (Tu et al. 2001), which has been documented in at least one case of human disease (Tu et al. 2004). The most recently described of these species, *C. insulaenigrae*, has been isolated from several marine mammals: common seals and a harbor porpoise in Scotland (Foster et al. 2004) and northern elephant seals in California (Stoddard et al. 2007). It too has been observed in a case of invasive human disease (Chua et al. 2007). Various *Campylobacter* species are routinely isolated from sewage

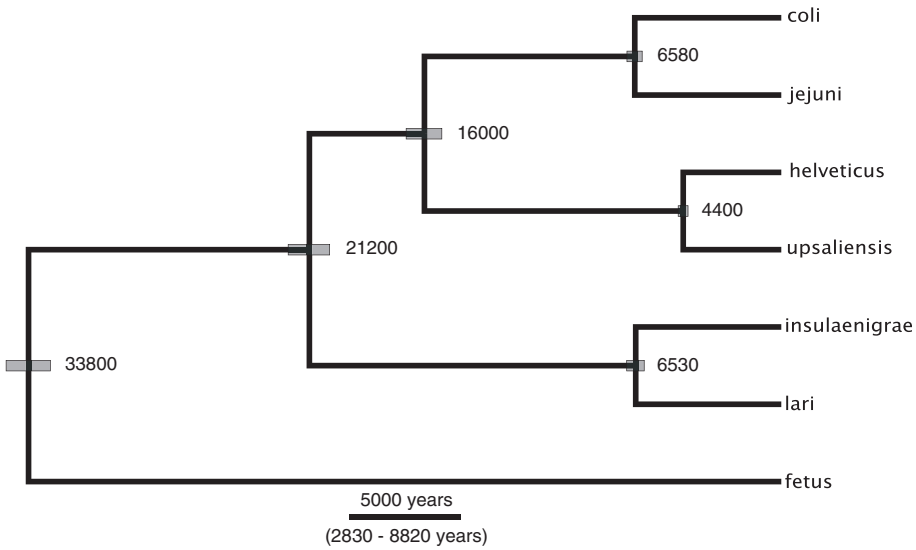

FIG. 4.—Phylogeny of the genus *Campylobacter*. Nodes are labeled with estimated divergence times using BEAST. Error bars associated with each node indicate relative uncertainty in node height. Uncertainty due to calibration of the molecular clock is represented by a 95% CI below the scale bar. Posterior uncertainty in the tree topology was negligible. The scale bar was calibrated from intraspecific variation in *Campylobacter jejuni*. For alternative scales, see table 7.

and environmental sources, including fresh and marine water (Jones 2001).

For each species, a typical isolate was chosen (see supplementary methods, Supplementary Material online). As no recombination was detected between these sequences ($P = 0.47$), we constructed a phylogeny using BEAST (Drummond et al. 2002), which we calibrated using our population genetic estimate of the synonymous mutation rate based on variation within *C. jejuni* (table 3). The standard method of dating recent bacterial evolution (Achtman et al. 2004; Roumagnac et al. 2006) is to calibrate the rate of sequence divergence relative to *E. coli* and *Salmonella typhimurium*, which Ochman and Wilson (1987) estimated to have split 120–160 Ma. They calculated a molecular clock of 1% per 50 My in the 16S rRNA gene. That would date the split between *C. jejuni* and *C. coli*, which differ by 0.4% on average (Gorkiewicz et al. 2003), to 10 Ma.

By our method, we obtained a vastly different estimate of 6,580 years ago (95% CI 3,580–12,400). The phylogeny in figure 4 is labeled with the inferred split times of all seven species. Two sources of uncertainty exist in the date estimates. The first source, which is modest in this case, is uncertainty in the split times relative to one another, indicated by the error bars associated with each node in the phylogeny. The second source, which dominates here, is uncertainty in the calibration of the molecular clock, which we represent as uncertainty in the scale bar of the phylogeny. Relative and total uncertainty in split times is detailed in table 5. There was negligible uncertainty in the tree topology.

Consistent with recent Neighbor-Joining (NJ) phylogenies based on the 16S rRNA, *rpoB*, and *groEL* genes (Kärenlampi et al. 2004; Korczak et al. 2006), we found that *C. jejuni* and *C. coli* are sister species, as are the pet-associated *C. helveticus* and *C. upsaliensis*, and the recently discovered *C. insulaenigrae* and *C. lari*. On the deeper structure, the NJ trees disagreed mutually, and with our Bayesian phylogeny, except on the observation that *C. fetus* is the most divergent of the seven species. The phylogeny we inferred is consistent with the observation of Fouts et al. (2005) that *C. upsaliensis* shares more protein identity (74.7%) with *C. jejuni* at the genomic level than does *C. lari* (68.9%).

Our method of calibration suggests that *Campylobacter* is speciating on the order of thousands, rather than millions of years. We dated the root of the tree to 33,800 years ago (95% CI 19,000–62,600). Using Yule's model (Yule 1924), a speciation rate of $\lambda = 0.0452$ per lineage per 1,000 years was inferred (table 6), which equates to an expected waiting time of 22,000 years between speciation events.

Calibration of the phylogeny is determined by the mutation rate, which we estimated at $\mu = 2.93 \times 10^{-2}$ per kb per year, slightly lower than the mutation rate within *C. jejuni*. The mutation rate is a function of the synonymous mutation rate (μ_S), which by our method was estimated from within *C. jejuni*, and the transition–transversion ratio (κ) and d_N/d_S ratio (ω). The latter two quantities were coestimated with the phylogeny but informed with prior estimates from within *C. jejuni*. As would be expected (Rocha et al.

Table 5
Phylogenetic Split Times in the Genus *Campylobacter*

Split	Point Estimate	95% CI	
		Relative Uncertainty	Total Uncertainty
coli–jejuni	6,580	6,240–6,930	3,580–12,400
helveticus– upsaliensis	4,400	4,190–4,630	2,400–8,290
insulaenigrae–lari	6,530	6,150–6,940	3,560–12,500
helveticus–jejuni	16,000	15,200–16,800	8,800–30,300
lari–jejuni	21,200	20,300–22,100	11,600–39,500
fetus–jejuni	33,800	32,800–34,800	19,000–62,600

Table 6
Evolutionary Parameters in the Genus *Campylobacter*

Parameter	Point Estimate	95% CI
μ	2.93×10^{-2}	1.60×10^{-2}–4.99×10^{-2}
κ	1.81	1.51–2.15
ω	0.0120	0.00961–0.0152
λ	0.0452	0.0149–0.120

Table 7
Alternative Scale Bars for *Campylobacter* Phylogeny

Method	Scale (Years)	95% CI
Intraspecific[a]	5,000	2,830–8,820
Empirical[b]	42,200	2,690–661,000
Ochman–Wilson[c]	7,600,000	Not quantified
Coalescent[d]	$23.6N_eg$	14.1 N_eg–39.6 N_eg

[a] Based on μ_S from table 3.
[b] Based on μ_S calculated from μ_0 in table 4.
[c] Based on Dingle et al. (2005).
[d] Based on θ_S from table 1.

2006), the d_N/d_S ratio was lower and in this case significantly lower ($\omega = 0.0120$), between *Campylobacter* species than within *C. jejuni*, which accounts for the lower total mutation rate. Curiously, we obtained a transition–transversion ratio 10-fold lower ($\kappa = 1.81$) between *Campylobacter* species than within *C. jejuni*.

To compare the intraspecific, population genetics calibration with other methods of calibration, we offer alternative scales in table 7 for the scale bar in figure 4. The empirical method utilizes the neutral molecular clock estimate from table 4, which is based on in vitro doubling times and experimental estimates of the per-generation mutation rate. Using the empirical approach, the length of the scale bar is 42,200 years with a wide CI (2,690–661,000) that subsumes the CI for the intraspecific method. The Ochman and Wilson (1987) method has been used in several recent publications (e.g., Achtman et al. 2004, Roumagnac et al. 2006) to calibrate the timescale of bacterial evolution and makes the length of the scale bar 7.6 My. The quantification of uncertainty in the original paper did not incorporate all sources of error, and so we do not put forward a CI. Finally, the coalescent approach uses $\theta_S/2$ for the synonymous rate of molecular change, which yields date estimates in coalescent time units of N_eg years. The T_{MRCA} for *C. jejuni* occurred around $2N_eg$ years ago, so this provides a relative means of dating the phylogeny. The length of the scale bar is $23.6N_eg$, with a CI of 14.1–$39.6N_eg$. That makes the *C. jejuni–C. coli* split around 17 times more ancient than the MRCA of *C. jejuni*.

Epidemiological Clustering Might Distort Calibration

In light of the 1,000-fold discrepancy between our intraspecific method of calibrating the rate of *Campylobacter* evolution and conventional estimates, we performed further statistical tests of robustness. We defer biological considerations of the plausibility of our estimates to the discussion. The formal hypothesis test we performed earlier strongly supported a model in which *C. jejuni* is measurably evolving, over one in which it is not. When above 1, the Bayes factor supports the measurably evolving model, and when below 1, it supports the simpler model in which all sequences were treated as if they were sampled simultaneously. We obtained a Bayes factor of 3.0×10^{20}, which is much greater than 1. However, for completeness, we performed a permutation test in which the sampling times of the sequences were randomized 100 times, and the Bayes factor recomputed in each case in order to obtain a reference distribution. No Bayes factor even remotely as large as 3.0×10^{20} was observed during the permutations, confirming that a signal consistent with measurable evolution does indeed exist.

It is possible that epidemiological clustering over time could cause an artifactual signal of ongoing evolution. For example, successive epidemics of genetically related *C. jejuni* sweeping through the human population could cause a correlation between sampling time and genetic distance: Organisms sampled closer together in time would be genetically more similar, mimicking the pattern expected in a measurably evolving population. There was no obvious pattern of recurrent epidemics in the data—indeed *C. jejuni* generally occurs sporadically (Wilson et al. 2008)—but in any event it would be difficult to distinguish from ongoing evolution. In an attempt to eliminate subtle epidemiological clustering, we thinned the data set so that no two cases occurred fewer than seven days apart, leaving 116 sequences. Having removed 90% of the sequences, we obtained a much smaller Bayes factor in favor of the measurably evolving model of 33.1. We repeated the permutation test as before. This time, out of 100 permutations, a Bayes factor larger than 33.1 was obtained three times. This suggests the measurably evolving model is still preferred, but much less so than before. It remains unclear whether by thinning the data, a large reduction in the Bayes factor was obtained because epidemiological clustering was causing an artifactual signal of ongoing evolution or because removing 90% of the sequences weakens inferential power.

Discussion

A Neolithic Origin for *C. jejuni*

Starkly at variance with conventional estimates of the timescale of bacteria evolution, how plausible is the molecular clock we estimated from within-species variation? Our estimate of the *C. jejuni–C. coli* split time coincides with the Neolithic domestication of a wide variety of animal and plant species, a time of broad cultural transition known as the Neolithic revolution, which began around 10,000 years ago. In particular, a date of 6,580 years ago coincides roughly with the spread of domestic pigs from the Near East and the first known domestication of wild boar of European descent in the Paris Basin, in the early fourth millennium BC (Larson et al. 2005, 2007). The relevance of this is the particular association of *C. coli* with domestic pigs (Dingle et al. 2005), which invites speculation as to the role of the advent of the domestic pig in driving the speciation of *C. jejuni* or *C. coli* from their common ancestor.

Interestingly, evidence is accumulating (Mira et al. 2006) that the Neolithic revolution played an important role

in creating new niches for a variety of pathogens of humans and their domesticated species, perhaps as a result of changes in agricultural practice or the advent of animal domestication. The genomes of pathogens such as *Bordetella pertussis* (whooping cough in humans), *Pseudomonas syringae* pathovar *tomato* (bacterial speck on tomatoes), and *Burkholderia mallei* (glanders in horses) exhibit a proliferation of insertion sequences connected with niche specialization and subsequent genome reduction/corruption. However, *C. jejuni* is unusual in harboring virtually no insertion sequences (Parkhill et al. 2000), and both it and *C. coli* retain wide host ranges.

No Resolution for the Bacterial Molecular Clock

Our synonymous rate estimate of $\mu_S = 2.79 \times 10^{-2}$ per kb per year is considerably at odds with traditional estimates of the bacterial molecular clock (Ochman and Wilson 1987), which would date the *C. jejuni–C. coli* split at 10 Ma. There are of course a number of reasons to exercise caution. Our Bayesian approach accounts for all sources of evolutionary uncertainty under the model, but uncertainty in the fundamental model assumptions is not so easily quantified. Extrapolation is the most dangerous form of statistical prediction, so the assumption that mutation rates are constant within and between species is a major one. It is widely appreciated that the long-term molecular clock is slower than short term (Rocha et al. 2006), but this is generally thought to be accounted for by the delayed effects of purifying selection. Although we attempted to abate the problem by assuming only that the synonymous rate is constant, we saw that not just the d_N/d_S ratio but also the transition–transversion ratio was significantly lower between species than within *C. jejuni*. Secondly, we demonstrated that a signal exists indicating that *C. jejuni* is measurably evolving on the timescale of our sample, but we cannot definitively rule out an artifactual association between sampling time and genetic similarity. For example, concurrent epidemic clusters sweeping through the population might generate such an association, although there was no evidence for it.

Even so, a rapid bacterial molecular clock cannot be dismissed out of hand. Indeed, two other attempts to calibrate the rate of bacterial evolution from intraspecific genetic variation have yielded similar estimates. Pérez-Losada et al. (2007) estimated a mutation rate of 4.58×10^{-2} kb^{-1} year^{-1} for housekeeping genes in *Neisseria gonorrhoeae* based on surveys of gonorrhoea patients in Baltimore, MD, between 1991 and 2005. Falush et al. (2001) used sequential biopsies from carriers of *Helicobacter pylori* in New Orleans, LA, and Colombia to estimate that the rate of mutation in housekeeping genes may have been as high as 4.1×10^{-2} kb^{-1} year^{-1}.

Traditional rate estimates are themselves at odds with empirical mutation rate estimates (Lenski et al. 2003; Ochman 2003), as we found here. Per-generation mutation rates estimated in the laboratory are several orders of magnitude higher than expected from Ochman and Wilson's molecular clock, even allowing for natural selection. Furthermore, certain commensal and pathogenic bacteria whose divergence can be dated via their host species exhibit 16S rRNA genetic diversity that is higher than expected from a rate of 1% per 50 My (Mira et al. 2006). Our own empirical estimates of the rate of molecular evolution were closer to our population genetic estimates than to the Ochman–Wilson estimates, differing by a factor of 10 in their point estimates and overlapping partially in their CIs. That discrepancy could be explained if Drake's method of mutation rate estimation (Drake 1991, Drake et al. 1998) gives a conservative estimate. Drake's method makes many assumptions and does not take account of lineage-specific knowledge. For example, *C. jejuni* lacks a number of genes responsible for DNA repair (Parkhill et al. 2000), which could underestimate the mutation rate sufficiently to explain the discrepancy between empirical and population genetics estimates. What we can refute is the suggestion (Tu et al. 2001) that mammal and reptile-associated *C. fetus* genotypes diverged 200 Ma, prior to the origin of mammals. This hypothesis is not even supported by the Ochman–Wilson method of calibration, which puts the *C. jejuni–C. fetus* split, the root of our phylogeny, at a mere 51.4 Ma.

Conclusion

Patterns of genetic variation within and between species provide an, albeit corrupted, document of evolutionary history. The picture of evolutionary history painted by genetic diversity in *C. jejuni* is one of a dynamically evolving species shaped by frequent recombination and intense purifying selection. *Campylobacter jejuni* is highly adaptable by virtue of its large effective population size, which counters a mutation rate that is low in comparison with viral pathogens (Drake et al. 1998). Together with routine cross-species gene flow, our analysis reveals a pathogen with the potential to adapt rapidly to changes in selection pressure.

Comparison with gene sequences from related species shows that *C. jejuni* may have evolved recently, within the past 12,000 years and possibly in response to changes in agricultural practice and the advent of animal domestication coinciding with the Neolithic revolution. *Campylobacter* is a dynamic genus in which species are no boundary to gene flow. Host ranges are wide and overlapping, and zoonosis between host species is common.

We believe that studies of ancient DNA offer the best prospect for accurately calibrating recent evolution in bacteria, although progress so far has been limited (Willerslev et al. 2004; Barnes and Thomas 2006). We have shown that large within-species samples can be analyzed with complex evolutionary models that incorporate recombination, without resorting to phylogenetic tree building. By integrating microevolutionary (population genetic) and macroevolutionary (phylogenetic) approaches in a Bayesian manner, we were able to quantify cumulative evolutionary uncertainty and perform detailed evolutionary inference. Certain gaps in our knowledge have been highlighted, including the generation length of pathogens in vivo and the per-generation rate of mutation in all but a few species. Finally, we have proposed that *Campylobacter* may be evolving on

a timescale of thousands, rather than millions of years. Our results contradict the traditional view of the rate of evolution in bacteria, but at the very least, they demand a re-evaluation of molecular clock estimates that are widely in use 20 years on.

Supplementary Material

Supplementary methods and supplementary figure S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors would like to thank David Balding, Mark Beaumont, Bob Griffiths, Rosalind Harding, Martin Maiden, Noel McCarthy, Gil McVean, Julian Parkhill, Andrew Rambaut, Roisin Ure, and Ian Wilson for advice and useful discussion. This work is part of the Veterinary Training Research Initiative, jointly funded by the Higher Education Funding Council of England and the Department for Environment, Food, and Rural Affairs. Funding was also received from the Engineering and Physical Sciences Research Council.

Literature Cited

Achtman M, Morelli G, Zhu P, et al. (17 co-authors). 2004. Microevolution and history of the plague bacillus, *Yersinia pestis*. Proc Natl Acad Sci USA. 101:17837–17842.

Adak GK, Meakins SM, Yip H, Lopman BA, O'Brien SJ. 2005. Disease risks from foods, England and Wales, 1996–2000. Emerg Infect Dis. 11:365–372.

Barnes I, Thomas MG. 2006. Evaluating bacterial pathogen DNA preservation in museum osteological collections. Proc R Soc Lond B. 273:645–653.

Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. Genetics. 162: 2025–2035.

Chua K, Gürtler V, Montgomery J, Fraenkel M, Mayall BC, Grayson ML. 2007. *Campylobacter insulaenigrae* causing septicaemia and enteritis. J Med Microbiol. 56:1565–1567.

de Boer P, Wagenaar JA, Achterberg RP, van Putten JPM, Schouls LM, Duim B. 2002. Generation of *Campylobacter jejuni* genetic diversity in vivo. Mol Microbiol. 44:351–359.

Dingle KE, Colles FM, Falush D, Maiden MCJ. 2005. Sequence typing and comparison of population biology of *Campylobacter coli* and *Campylobacter jejuni*. J Clin Microbiol. 43: 340–347.

Dingle KE, Colles FM, Wareing DRA, Maiden MCJ, Ure MCJ, Maiden R, Fox AJ, Bolton FE, Bootsma HJ, Willems RJ, Urwin R, Maiden MC. 2001. Multilocus sequence typing system for *Campylobacter jejuni*. J Clin Microbiol. 39:14–23.

Drake JW. 1991. A constant rate of spontaneous mutation in DNA-based microbes. Proc Natl Acad Sci USA. 88:7160–7164.

Drake JW, Charlesworth B, Charlesworth D, Crow JF. 1998. Rates of spontaneous mutation. Genetics. 148:1667–1686.

Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics. 161:1307–1320.

Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG. 2003. Measurably evolving populations. Trends Ecol Evol. 18:481–488.

Falush D, Kraft C, Taylor NS, Correa P, Fox JG, Achtman M, Suerbaum S. 2001. Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. Proc Natl Acad Sci USA. 98:15056–15061.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure from multilocus genotype data: linked loci and correlated allele frequencies. Genetics. 164:1567–1587.

Fearnhead P. 2008. Computational methods for complex stochastic systems: a review of some alternatives to MCMC. Stat Comput. 18:151–171.

Fearnhead P, Smith NGC, Barrigas M, Fox A, French N. 2005. Analysis of recombination in *Campylobacter jejuni* from MLST population data. J Mol Evol. 61:333–340.

Foster G, Holmes B, Lawsono PA, Thorne P, Byrer DE, Ross HM, Xerry J, Thompson PM, Collins MD. 2004. *Campylobacter insulaenigrae* sp. nov., isolated from marine mammals. Int J Syst Evol Microbiol. 54:2369–2373.

Fouts DE, Mongodin EF, Mandrell RE, et al. (21 co-authors). 2005. Major structural differences and novel potential virulence mechanisms from the genomes of multiple *Campylobacter* species. PLoS Biol. 3:e15.

Gelman A, Carlin JB, Stern HS, Rubin DB. 2003. Bayesian data analysis, 2nd ed. Boca Raton (FL): Chapman Hall.

Gillespie IA, O'Brien SJ, Frost JA, Adak GK, Horby P, Swan AV, Painter MJ, Neal KR. 2002. A case–case comparison of *Campylobacter coli* and *Campylobacter jejuni* infection: a tool for generating hypotheses. Emerg Infect Dis. 8:937–942.

Gorkiewicz G, Feierl G, Schober C, Dieber F, Köfer J, Zechner R, Zechner EL. 2003. Species-specific identification of Campylobacters by partial 16S rRNA gene sequencing. J Clin Microbiol. 41:2537–2546.

Humphrey T, O'Brien SJ, Madsen M. 2007. Campylobacters as zoonotic pathogens: a food production perspective. Int J Food Microbiol. 117:237–257.

Jackson RJ, Elvers KT, Lee LJ, Gidley MD, Wainwright LM, Lightfoot J, Park SF, Poole RK. 2007. Oxygen reactivity of both respiratory oxidases in *Campylobacter jejuni*: the *cydAB* genes encode a cyanide-resistant, low-affinity oxidase that is not of the cytochrome *bd* type. J Bacteriol. 189: 1604–1615.

Jones K. 2001. Campylobacters in water, sewage and the environment. J Appl Microbiol. 90:68S–79S.

Kärenlampi RI, Tolvanen TP, Hänninen M-L. 2004. Phylogenetic analysis and PCR-restriction fragment length polymorphism identification of *Campylobacter* species based on partial *groEL* gene sequences. J Clin Microbiol. 42:5731–5738.

Khanna MR, Bhavsar SP, Kapadnis BP. 2006. Effect of temperature on growth and chemotactic behaviour of *Campylobacter jejuni*. Lett Appl Microbiol. 43:84–90.

Kimura M. 1955. Stochastic processes and distributions of gene frequencies under natural selection. Cold Spring Harbor Symp. 20:33–53.

Kimura M, Crow JF. 1964. The number of alleles that can be maintained in a finite population. Genetics. 49:725–738.

Kingman JFC. 1982. On the genealogy of large populations. J Appl Prob. 19A:27–43.

Konkel ME, Christensen JE, Dhillon AS, Lane AB, Hare-Sanford R, Schaberg DM, Larson CL. 2007. *Campylobacter jejuni* strains compete for colonization in broiler chicks. Appl Environ Microbiol. 73:2297–2305.

Korczak BM, Stieber R, Emler S, et al. 2006. Genetic relatedness within the genus *Campylobacter* inferred from *rpoB* sequences. Int J Syst Evol Microbiol. 56:937–945.

Larson G, Albarella U, Dobney K, et al. (19 co-authors). 2007. Ancient DNA, pig domestication, and the spread of the

Neolithic into Europe. Proc Natl Acad Sci USA. 104: 15276–15281.
Larson G, Dobney K, Albarella U, et al. (13 co-authors). 2005. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. Science. 307:1618–1621.
Lenski RE, Winkworth CL, Riley MA. 2003. Rates of DNA sequence evolution in experimental populations of *Escherichia coli* during 20,000 generations. J Mol Evol. 56:498–508.
McCarthy ND, Colles FM, Dingle KE, Bagnall MC, Manning G, Maiden MC, Falush D. 2007. Host-associated genetic import in *Campylobacter jejuni*. Emerg Infect Dis. 13:267–272.
McVean GAT, Awadalla P, Fearnhead P. 2002. A coalescent method for detecting recombination from gene sequences. Genetics. 160:1231–1241.
Miller WG, On SLW, Wang G, Fontanoz S, Lastovica AJ, Mandrell RE. 2005. Extended multilocus sequence typing system for *Campylobacter coli*, *C. lari*, *C. upsaliensis*, and *C. helveticus*. J Clin Microbiol. 43:2315–2329.
Mira A, Pushker R, Rodríguez-Valera F. 2006. The Neolithic revolution of bacterial genomes. Trends Microbiol. 14:200–206.
Moore JE, Barton MD, Blair IS, et al. (18 co-authors). 2006. The epidemiology of antibiotic resistance in *Campylobacter*. Microbe Infec. 8:1955–1966.
Moran NA, Munson MA, Baumann P, Ishikawa H. 1993. A molecular clock in endosymbiotic bacteria is calibrated using insect hosts. Proc R Soc Lond B. 253:167–171.
Nielsen R, Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics. 148:929–936.
Ochman H. 2003. Neutral mutations and neutral substitutions in bacterial genomes. Mol Biol Evol. 20:2091–2096.
Ochman H, Wilson AC. 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J Mol Evol. 26:74–86.
Oyarzabal OA, Rad R, Backert S. 2007. Conjugative transfer of chromosomally encoded antibiotic resistance from *Helicobacter pylori* to *Campylobacter jejuni*. J Clin Microbiol. 45: 402–408.
Parkhill J, Wren BW, Mungall K, et al. (21 co-authors). 2000. The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. Nature. 403:665–668.
Pérez-Losada M, Browne EB, Madsen A, Wirth T, Viscidi RP, Crandall KA. 2006. Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. Infect Genet Evol. 6:97–112.
Pérez-Losada M, Crandall KA, Zenilman J, Viscidi RP. 2007. Temporal trends in gonococcal population genetics in a high prevalence urban community. Infect Genet Evol. 7:271–278.
Rocha EPC, Maynard Smith J, Hurst LD, Holden MTG, Cooper JE, Smith NH, Feil EJ. 2006. Comparisons of dN/dS are time dependent for closely related bacterial genomes. J Theor Biol. 239:226–235.
Rodrigo A, Felsenstein J. 1999. Coalescent approaches to HIV-1 population genetics. In: Crandall K, editor. Molecular evolution of HIV. Baltimore (MD): John Hopkins University Press. p. 233–272.
Roumagnac P, Weill F-X, Dolecek C, et al. (11 co-authors). 2006. Evolutionary history of *Salmonella* typhi. Science. 314: 1301–1304.
Sheppard SK, McCarthy ND, Falush D, Maiden MCJ. 2008. Convergence of *Campylobacter* species: implications for bacterial evolution. Science. 320:237–239.
Stoddard RA, Miller WG, Foley JE, Lawrence J, Gulland FMD, Conrad PA, Byrne BA. 2007. *Campylobacter insulaenigrae* isolates from northern elephant seals (*Mirounga angustirostris*) in California. Appl Environ Microbiol. 73:1729–1735.
Tu Z-C, Dewhirst FE, Blaser MJ. 2001. Evidence that the *Campylobacter fetus sap* locus is an ancient genomic constituent with origins before mammals and reptiles diverged. Infect Immun. 69:2237–2244.
Tu Z-C, Zeitlin G, Gagner J-P, Keo T, Hanna BA, Blaser MJ. 2004. *Campylobacter fetus* of reptile origin as a human pathogen. J Clin Microbiol. 42:4405–4407.
van Bergen MAP, Dingle KE, Maiden MCJ, Newell DG, van der Graaf-Van Bloois L, van Putten JP, Wagenaar JA. 2005. Clonal nature of *Campylobacter fetus* as defined by multilocus sequence typing. J Clin Microbiol. 43:5888–5898.
Velayudhan J, Kelly DJ. 2002. Analysis of gluconeogenic and anaplerotic enzymes in *Campylobacter jejuni*: an essential role for phosphoenolpyruvate carboxykinase. Microbiology. 148:685–694.
Willerslev E, Hansen A, Rønn R, Brand TB, Barnes I, Wiuf C, Gilichinsky D, Mitchell D, Cooper A. 2004. Long-term persistence of bacterial DNA. Curr Biol. 14:R9–R10.
Wilson DJ, Gabriel E, Leatherbarrow AJH, Cheesbrough J, Gee S, Bolton E, Fox A, Hart CA, Diggle PJ, Fearnhead F. 2008. PLoS Genet. 4:e1000203.
Wiuf C, Hein J. 2000. The coalescent with gene conversion. Genetics. 155:451–462.
Yule GU. 1924. A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis. Philos Trans Roy Soc Lond Ser B. 213:21–87.
Zia S, Wareing D, Sutton C, et al. 2003. Health problems following *Campylobacter jejuni* enteritis in a Lancashire population. Rheumatology. 42:1083–1088.

Rasmus Nielsen, Associate Editor

Accepted November 6, 2008