

Genome-wide association study identifies vitamin B₅ biosynthesis as a host specificity factor in *Campylobacter*

Samuel K. Sheppard^{a,b,1}, Xavier Didelot^c, Guillaume Meric^b, Alicia Torralbo^d, Keith A. Jolley^a, David J. Kelly^e, Stephen D. Bentley^{f,g}, Martin C. J. Maiden^a, Julian Parkhill^f, and Daniel Falush^h

^aDepartment of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom; ^bInstitute of Life Science, College of Medicine, Swansea University, Swansea SA2 8PP, United Kingdom; ^cSchool of Public Health, St. Mary's Campus, Imperial College London, London SW7 2AZ, United Kingdom; ^dDepartment of Animal Health, University of Cordoba, 14071 Cordoba, Spain; ^eDepartment of Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2TN, United Kingdom; ^fWellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SA, United Kingdom; ^gDepartment of Medicine, University of Cambridge, Addenbrookes Hospital, Cambridge CB2 0SP, United Kingdom; and ^hMax Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany

Edited by W. Ford Doolittle, Dalhousie University, Halifax, Canada, and approved June 3, 2013 (received for review March 22, 2013)

Genome-wide association studies have the potential to identify causal genetic factors underlying important phenotypes but have rarely been performed in bacteria. We present an association mapping method that takes into account the clonal population structure of bacteria and is applicable to both core and accessory genome variation. *Campylobacter* is a common cause of human gastroenteritis as a consequence of its proliferation in multiple farm animal species and its transmission via contaminated meat and poultry. We applied our association mapping method to identify the factors responsible for adaptation to cattle and chickens among 192 *Campylobacter* isolates from these and other host sources. Phylogenetic analysis implied frequent host switching but also showed that some lineages were strongly associated with particular hosts. A seven-gene region with a host association signal was found. Genes in this region were almost universally present in cattle but were frequently absent in isolates from chickens and wild birds. Three of the seven genes encoded vitamin B₅ biosynthesis. We found that isolates from cattle were better able to grow in vitamin B₅-depleted media and propose that this difference may be an adaptation to host diet.

evolution | genomics | host adaptation | transmission ecology

Colonization of multiple host species increases the number of transmission opportunities for animal pathogens and symbionts but depends on making rapid adjustments to each new host (1). For organisms such as *Campylobacter*, relatively small genome size (1.6 Mb) limits the phenotypic flexibility of each bacterium. Single clones can multiply to large numbers within hosts, and genetic variation arising among these bacteria increases the range of available phenotypes. This might allow a bacterial lineage to passage successfully through multiple hosts by repeatedly evolving host adaptive traits.

Experimental work has shown that a large proportion of adaptations to new environments incur an equal or greater cost in other environments (2). This cost of adaptation might make a strategy of continuous evolution unstable by causing a progressive loss of fitness in the course of repeated host switching. Three factors that could reduce this cost of readaptation are canalization of genetic change via contingency loci (3, 4); coordinated genetic regulation of host-specific factors (5, 6); and import of DNA by recombination from other, already adapted, lineages in each new host species (7). The relative importance of these mechanisms for host specificity in *Campylobacter* remains unknown.

Campylobacter jejuni and *Campylobacter coli* are common components of the gut microbiota in numerous wild and domesticated animal species, as well as, together, being one of the most common causes of food poisoning in humans. The characterization of large numbers of *C. jejuni* and *C. coli* isolates from diverse

sources and locations by multilocus sequence typing (MLST) has shown that there is genetic differentiation among sequence types (STs) associated with different hosts (8). Among wild birds, specific bird species most often harbor their own *Campylobacter* lineages (8, 9). However, in agricultural animals, although there are host-associated lineages that are largely restricted to chickens or to ruminants, some of the most abundant lineages are found at high frequencies in chickens, cattle, and humans with food poisoning (8). This multihost lifestyle is curious because of the challenges of colonizing organisms with such distinct gastrointestinal tracts, diets, immune systems, and body temperatures.

Here, we investigated the genetic basis of host specificity by analyzing the genome sequences of 192 isolates from cattle, chickens, clinical samples, and other sources. These included isolates from single-host lineages and multiple isolates from the host generalist ST-21 (*C. jejuni*), ST-45 (*C. jejuni*), and ST-828 (*C. coli*) clonal complexes (Table S1). We used phylogenetic analysis to investigate host switching and then sought to identify genetic elements that showed a stronger association with chicken or ruminants than expected based on the phylogeny. These elements represent candidate host adaptation elements. We identified one substantial cattle-associated region and demonstrate experimentally that it confers a phenotype (vitamin B₅ biosynthesis) likely to aid in colonization of cattle.

Results

Clonal complexes, identified using MLST data, are groups of isolates that have identical sequences at most of or all the seven genotyped loci. Now that whole genomes are becoming available for large numbers of bacteria, it is possible to establish the accuracy of clonal complex designations in identifying clonal lineages. Phylogenetic analysis of *C. jejuni* genomes supports most of the clonal complex groupings identified by MLST and provides new insights into their relationships (Fig. 1A). For example, isolates from the chicken-associated ST-354, ST-443, ST-353, and ST-257 complexes (8) each form distinct clades in the tree

Author contributions: S.K.S., X.D., and D.F. designed research; S.K.S., X.D., G.M., A.T., S.D.B., and J.P. performed research; S.K.S., X.D., K.A.J., M.C.J.M., and J.P. contributed new reagents/analytic tools; S.K.S., X.D., G.M., D.J.K., and D.F. analyzed data; and S.K.S., X.D., and D.F. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Genome sequence data for isolates that were sequenced in this study have been deposited with Dryad, <http://datadryad.org/>. The data will be available in 48 h from the date of resubmission with the following doi:10.5061/dryad.28n35. Data will also be available via PubMLST, <http://pubmlst.org/campylobacter/>.

¹To whom correspondence should be addressed. E-mail: s.k.sheppard@swansea.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1305559110/-DCSupplemental.

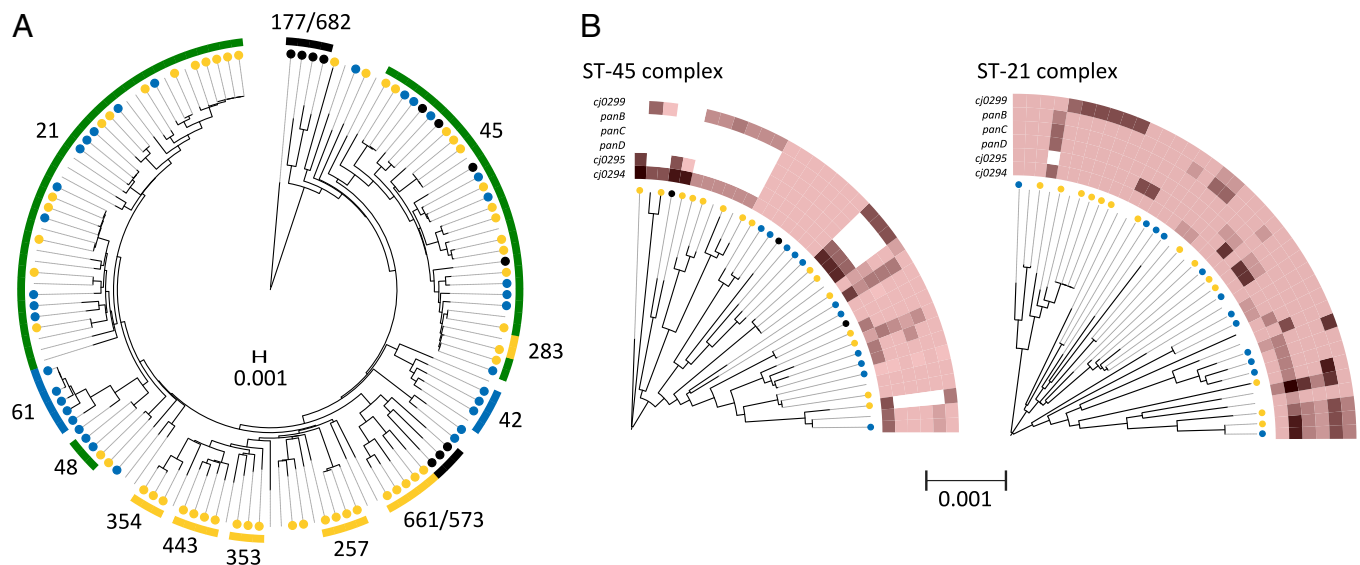


Fig. 1. Genetic structure of *C. jejuni* isolates from different hosts. (A) Neighbor-joining tree of all isolates based on 1.53-Mb concatenated sequences (165,206 variable sites) of 1,623 loci in the NCTC11168 *C. jejuni* isolate genome. Host origin is indicated for chickens (yellow), cattle (blue), and wild birds and the environment (black). Other isolates are primarily from human infections. Clonal complex designations based on MLST are labeled around the tree. Host associations, based on the 2,764 isolates in the pubMLST database (pubMLST.org), are indicated in the halo using the same color scheme, with generalist lineages shown in green. (B) Tree of the ST-45 and ST-21 clonal complex isolates only, estimated using ClonalFrame. Allelic variation for six genes in the host-associated region is shown at the end of each branch. Each shade of red represents a unique allele at that locus. White denotes the absence or truncation of the gene.

and, together with a handful of other isolates, form a single chicken-associated supercomplex. Isolates from other chicken-associated clonal complexes, such as the ST-661 complex, branch elsewhere on the tree. There are also two cattle-associated lineages corresponding to the ST-61 and ST-42 complexes. In addition, there are lineages with levels of genetic variation comparable to those of the host-associated groups: specifically, the ST-21 and ST-45 complexes containing a mixture of isolates from cattle, chickens, and wild birds.

To investigate relatedness within the multihost clonal complexes, we used ClonalFrame (10), which estimates clonal relationships allowing for the effect of recombination. Chicken and cattle isolates are found in multiple places on these trees (Fig. 1B), which implies relatively recent host switching, although there is one deeper branch of the ST-45 complex that may be chicken-associated.

To investigate the evolution of host specialization, we developed a genome-wide association mapping approach that identifies 30-bp DNA sequences (words) associated with colonization of particular species. The method identifies words that are more strongly associated with a particular host than would be expected based on neutral patterns of evolution, given the clonal relationships of the bacteria in the sample and their distribution among hosts. An attractive feature of this method is that it has the potential to identify host-adapting evolutionary events, including point mutation, homologous recombination, and lateral gene transfer.

In order for the method to have statistical power to detect associations, the dataset needs to contain multiple host switches within it. Furthermore, the presence of multiple isolates from the same host and clonal background (e.g., the chicken supercomplex) can reduce power because these isolates will share large amounts of DNA sequence due to clonal inheritance that do not confer host adaptive traits. We initially applied our method to 29 genome sequences from the ST-45 clonal complex, which are commonly found in wild and domesticated bird and mammal species. We included two genomes belonging to the ST-283 complex, which clusters within the ST-45 complex. We chose this lineage because it met the requirement of multiple host switches,

and we reasoned that the low overall genetic variation within the lineage should enhance the power to detect adaptive events.

A total of 9,034 30-bp words were identified that were significantly associated with either chicken or cattle hosts. Of these, 8,999 mapped to 97 genes (Table S2) in the annotated genome of isolate NCTC11168 (11) in 10 genome locations (Fig. 2A). The association signal was replicated by determining the number of these words in the genomes of 161 *C. jejuni* and *C. coli* isolates from outside the ST-45 clonal complex. The pattern of host association varied among regions in the replication (Table S3). The most concordant signal, where the same pattern of host association occurred in the ST-45 complex isolates and in the remaining genomes, was among 7,307 cattle-associated words that came from 7 adjacent genes within a 5-Kb region (Table 1). We therefore focused on the evolutionary history of this region to investigate the events that generated the association signal.

The pattern of gene presence varied for the six genes at the center of this region. The number of isolates from which they were completely absent varied from 3 isolates for *surE* to 78 and 98 isolates for *Cj0299* and *Cj0295*, respectively. The *panBCD* genes were absent from between 34 and 39 isolates. Although the genes showed different patterns of presence and absence, they were all more common in cattle isolates than in chicken isolates. For example, in the ST-45 complex, the *panBCD* genes had been gained or lost three times (Fig. 1) and were not present in the two largest clades of the tree, where cattle isolates were absent. In contrast, the *panBCD* genes were present in all the ST-21 complex isolates examined.

The host association signal was the result of two forms of genetic variation. First, with the exception of *surE*, which had fewer host-associated words, all the genes in the major host-associated region were more likely to be absent in isolates from chickens. Second, even where genes were present, different sequences were found in cattle and chicken isolates. Alleles at these loci tended to show less homologous sequence variation in cattle isolates compared with chicken isolates, and as a result, cattle-associated alleles gave the strongest signals of host association. It is also notable that the

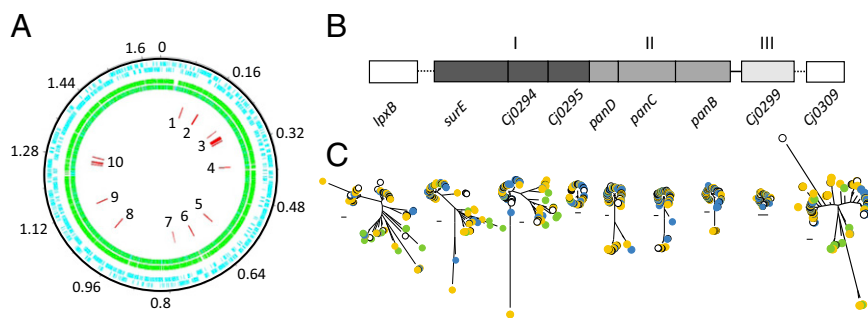


Fig. 2. Host-associated genomic regions. (A) Distribution of 10 locations containing 8,538 host-associated words for which homologs were found in reference genome NCTC11168, visualized using Artemis (12) and DNAPlotter (13). The red lines indicate host-associated words, the cyan rings show the protein coding sequences on the forward and reverse strands, and the green rings indicate other annotated features using default settings. (B) Schematic representation of seven adjacent genes within a 5-Kb region at location 3, to which 7,307 cattle-associated words mapped, and example flanking genes for comparison. Three regions (I–III) were identified based on host differences in word presence (Fig. S1). (C) Neighbor-joining trees of these seven genes in isolates from cattle (blue), chickens (yellow), and wild birds/environmental waters (green). The scale bars represent a genetic distance of 0.005.

frequency of both types of host-associated words was even lower in isolates from wild birds than in those from chickens (Fig. S1).

The majority of the cattle-associated words mapped to the three genes from the pantothenate (vitamin B₅) biosynthesis pathway (*panBCD*). To test if cattle-associated isolates had a greater capacity to grow in a vitamin B₅-depleted environment, we conducted in vitro growth experiments. This showed that isolates from cattle grew better, on average, in a low vitamin B₅ environment than isolates from chickens (Fig. 3) and that this was due, at least in part, to the presence of the *panBCD* genes in a higher proportion of cattle isolates (Fig. S2).

The gene *Cj0299* had exceptionally low sequence variation relative to other commonly present genes (Fig. S3), with only seven polymorphic sites among 774 nucleotides. This gene, which encodes a characterized enzyme called *bla*_{OXA-61} conferring resistance to β-lactam antibiotics (12), was found at the highest frequency in cattle isolates and was rarest in isolates from wild birds (Fig. S1).

Discussion

Our analysis provides insight into the tempo of adaptation of *Campylobacter* to its host. Host specificity can be a stable trait in some lineages. For example, specialization to chicken hosts presumably evolved in the common ancestor of the chicken super-complex (Fig. 1A). There are also other less diverse chicken- and cattle-associated lineages. In addition, there are generalist lineages that are abundant in cattle, chickens, and other hosts. Within these lineages, there are sublineages with evidence of host association, suggesting that host specificity can also evolve on a much more rapid time scale.

Using the association mapping method described here, we have shown that one factor driving rapid host adaptation is gain and loss of the *panBCD* genes encoding the vitamin B₅ biosynthesis pathway, which has already been highlighted as especially variable in a previous study (13). Vitamin B₅ is abundant in cereals and grains but is found in a very low concentration in grasses (14), which, respectively, constitute the main diets of chicken and cattle. This makes it likely that *Campylobacter* needs to produce the vitamin itself to persist in cattle. We found that the genes are almost universally present in isolates from cattle.

Our methods for detecting association, like those used in humans (15, 16), are influenced by genetic linkage. Linkage affects association signals because the recombination events that lead to the gain or loss of host-adaptive elements can also affect the neighboring sequence. The situation is complicated by the possibility that gene order also evolves in response to selection for linkage (17), leading to elements that give advantages in the same or overlapping environments being inherited together, for example, in “pathogenicity islands” (18).

We found patterns of host association in the genes adjacent to *panBCD* similar to those in the genes themselves. Three upstream genes (*surE*, *Cj0294*, and *Cj0295*) are also involved in biosynthesis. Immediately downstream is an ampicillin resistance-associated gene, *bla*_{OXA-61} (12), which is also associated with cattle. The gene is notable for being the least diverse of the common genes in the *Campylobacter* genome (Fig. S3), implying that it has recently spread among the diverse *C. coli* and *C. jejuni* lineages in our sample. It is intriguing that these seven adjacent genes appear to code for two functions related to the challenges of the agricultural niche, namely, antibiotics and persistence in ruminants. More data will be needed to determine the extent to which selection is acting

Table 1. Genes with cattle associated alleles

Gene name*	Description
<i>surE</i> (<i>Cj0293</i>)	Stationary phase survival protein; catalyzes the conversion of a phosphate monoester to an alcohol and a phosphate
<i>Cj0294</i>	Dinucleotide-using enzymes involved in molybdopterin and thiamine biosynthesis family 1
<i>Cj0295</i>	Putative acetyltransferase
<i>panD</i> (<i>Cj0296c</i>)	Aspartate α-decarboxylase; converts L-aspartate to β-alanine and provides the major route of β-alanine production in bacteria. β-alanine is essential for the biosynthesis of pantothenate (vitamin B ₅).
<i>panC</i> (<i>Cj0297c</i>)	Pantoate-β-alanine ligase; catalyzes the formation of (R)-pantothenate from pantoate and β-alanine
<i>panB</i> (<i>Cj0298c</i>)	3-Methyl-2-oxobutanoate hydroxymethyltransferase; catalyzes the formation of tetrahydrofolate and 2-dehydropantoate from 5,10-methylenetetrahydrofolate and 3-methyl-2-oxobutanoate
<i>Cj0299</i>	Putative periplasmic β-lactamase

*Gene names, numbering, and order are based on the *C. jejuni* strain 11168 annotation (11).

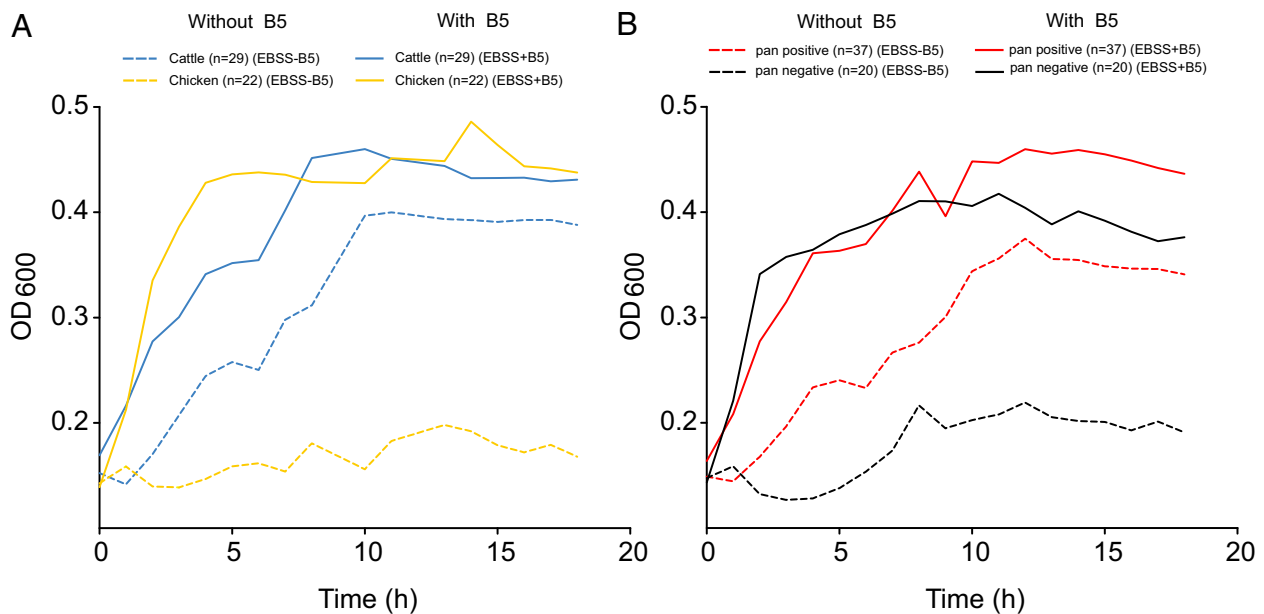


Fig. 3. Growth of *C. jejuni* in medium with and without pantothenic acid (vitamin B₅). Curves in *A* and *B* represent growth levels in vitro (OD₆₀₀) over a period of 18 h at 42 °C in microaerobic conditions in medium supplemented with vitamin B₅ (solid lines) and lacking vitamin B₅ (dashed lines). (*A*) Growth curves according to the host of origin, with cattle-associated isolates in blue ($n = 29$) and chicken-associated isolates in yellow ($n = 22$). (*B*) Growth curves according to the presence ($n = 37$, red lines) or absence ($n = 20$, black lines) of the *panBCD* operon in tested isolates.

specifically on *panBCD*, on *bla*_{OXA-61}, or on the combined effect of the genes together.

Our results suggest that host generalism in agricultural *Campylobacter* is possible because elements that are necessary for colonization of cattle (specifically, the genes responsible for vitamin B₅ biosynthesis) persist in some isolates in chickens. However, many details remain to be investigated, and it is not clear whether generalism is a stable strategy taking advantage of the large number of transmission opportunities in agriculture or whether it reflects insufficient time for well-adapted host specialist lineages to have evolved in this recent manmade environment.

As well as being found in many chicken isolates from host generalist lineages, the *panBCD* genes are found in isolates from chicken-associated lineages. It is not known how large a fitness cost, if any, these genes incur in environments in which vitamin B₅ is plentiful. Furthermore, part of the transmission ecology of chicken-associated *Campylobacter* lineages might depend on vitamin B₅ biosynthesis, for example, if spread is dependent on transition via cattle or other host species. This could compensate for the metabolic cost of maintaining these genes in the primary host.

Our study illustrates the potential of genome-wide association methods for studying rapidly evolving traits in bacteria. The molecular Koch's postulates, defined by Falkow (19), describe a set of experimental criteria that should be satisfied to show that a gene found in a pathogenic microorganism contributes to the disease caused by the pathogen. The postulates can also be applied to other microbial traits, such as host specificity. Genetic association study methods, such as those used here and elsewhere (20), will facilitate the identification of genetic elements likely to cause phenotypes of interest and provide targets for further laboratory investigation.

Materials and Methods

Isolates and Genome Sequencing. Isolates were chosen from multilocus sequence-typed collections that, together, contain >10,000 isolates (www.pubmlst.org/campylobacter) (21) to represent common STs found in clinical samples (feces) and those associated with various agricultural animal and wild bird species. In addition, multiple isolates from the ST-21 (*C. jejuni*), ST-45 (*C. jejuni*), and ST-828 (*C. coli*) clonal complexes were chosen from

different hosts to investigate the association of genomic elements with phenotype (host source). Details of all the isolates used, including genomes from published sources (22–26), such as the National Center for Biotechnology Information database, are provided in Table S1.

Campylobacter isolates were subcultured on Columbia Blood Agar, and plates were incubated in microaerobic conditions (5% CO₂, 5% O₂, 3% H₂, and 87% N₂) at 42 °C for 48 h. Single-colony cultures were harvested, cells were suspended in 125 μL of water in a 0.2-mL PCR tube, and extraction of genomic DNA was carried out using the QIAamp DNA Mini Kit (QIAGEN GmbH). The DNA was resuspended in 100–200 μL of the elution buffer supplied and stored at –20 °C.

Whole-genome sequencing was carried out using a multiplex sequencing approach on an Illumina Genome Analyzer using the standard indexing protocol. Fragmentation of 2 μg of genomic DNA was carried out by acoustic shearing using a Covaris E210 sonicator. Following enrichment for 200-bp fragments, DNA was end-repaired and cleaned using a 1:1 ratio of Ampure paramagnetic beads (Beckman Coulter, Inc.) to remove DNA <150 bp in length. A-tailing was carried out, and adapters were ligated, with each followed by additional DNA clean-up. An overlap extension PCR was carried out, using an Illumina 3 primer set, to introduce specific tag sequences between the sequencing and the flow-cell binding sites of an Illumina adapter. A final DNA clean-up was carried out before DNA quantification by quantitative PCR. Twelve separately tagged libraries were sequenced simultaneously in two lanes of an eight-channel Illumina Genome Analyzer II flow cell. The average overall output was 80 Mbp per isolate. High coverage short reads (25–50 bp) were assembled de novo using Velvet software (27) to produce contigs of up to 162 kb. The 80 genome sequences that are published in this paper are available in the Dryad repository (10.5061/dryad.28n35).

Background Population Genetic Structure. The publicly accessible Bacterial Isolate Genome Sequence Database (BIGSdb) (21) was used to store contiguous sequences and whole-genome data from the GenBank. The finished genome of isolate NCTC11168 (25, 28, 29) was used as a basis for defining locus designations and reference sequences for each of these that were stored in the database. The BLAST algorithm was used to scan other genomes for gene orthologs at these loci, defined as reciprocal best hits of a sequence with ≥70% nucleotide identity and a 50% difference in alignment length. MUSCLE software (30) was used to align gene orthologs on a gene-by-gene basis, and these data were concatenated into contiguous sequence for each isolate genome, including gaps for missing nucleotides (or entire genes). A whole-genome tree of alignments was reconstructed for the entire genus using MEGA (31) version 3.1 with the Kimura two-parameter model and neighbor-joining clustering. In addition, gene-by-gene alignments were

extracted from the BIGSdb (21) for genes present in all three major multihost lineages (ST-828, ST-21, and ST-45 clonal complexes). Genealogies for these alignments were estimated using ClonalFrame, a model-based approach to determining microevolution in bacteria (10). This program differentiates mutation and recombination events on each branch of the tree based on the density of polymorphisms. Clusters of polymorphisms are likely to have arisen from recombination, and scattered polymorphisms are likely to have arisen from mutation. The program was run with 50,000 burn-in iterations, followed by 50,000 sampling iterations. The consensus tree represents combined data from three independent runs with 75% consensus required for inference of relatedness. Recombination events were defined as sequences with a length of >50 bp with a probability of recombination $\geq 75\%$ over the length, reaching 95% in at least one site.

Association Mapping. One key aim of this study was to search for the genomic variants underlying preferential host colonization in *C. jejuni*. Such genome-wide association mapping has been extensively performed in humans (15, 16) but not in bacteria (32), which meant that methodology had to be developed to account for the specificities of bacterial population genetics. For each genome, the presence or absence of each unique "word" of 30 bp on the forward or reverse strand of any contig was recorded. Such short words are also known as "features," and they have been used previously to build alignment-free phylogenies (33, 34). Here, however, we wanted to measure the statistical association of the presence of a word with the host from which each genome was isolated. This approach presents the advantage of being scalable to large numbers of genomes, alignment-free, and able to encapsulate in the same framework any genomic variant, whether it is caused by point mutation, homologous recombination, or lateral gene transfer. One difficulty, however, is that naively measuring the association of word presence with the host (e.g., using a Fisher's exact test) would produce many spurious association due to the confounding effect of population structure. This is a well-known problem when performing association mapping in humans (35), but it is likely to be even more problematic in bacteria due to

their clonal mode of reproduction (32). To control for the population structure when measuring the association of a word with a host, we used a technique inspired from phylogenetic comparative methods through Monte Carlo simulations (36, 37). Based on the phylogeny reconstructed by ClonalFrame (10), words were simulated to evolve through a process of gain and loss along the branches of the phylogeny. The correlation of the real word was then compared with the distribution of the correlations of the simulated words to produce a phylogenetically correct *P* value (38). To account for multiple testing, only words with a *P* value below 10^{-4} were considered significant. The distribution of host-associated words for which homologs were found in reference genome NCTC11168 (11, 29) was visualized using Artemis (39), and DNAPlotter (40) was used to generate Fig. 2A. To test the sensitivity of the method to word size, we performed the same analysis with 20-bp and 40-bp words and obtained comparable results (Fig. S4).

Phenotypic Analyses. The ability to grow in controlled rich medium in presence or absence of vitamin B₅ was assessed in vitro. The synthetic medium used in this analysis was composed of 1× Earle's balanced salt solution (EBSS), 1× MEM essential amino acid solution, 1× MEM nonessential amino acids (catalog no. 24010-043; Invitrogen), 1 mM sodium pyruvate, 20 μM FeSO₄, and 2.1 μM D-pantothenic acid 1/2 Ca salts (catalog no. P2250; Sigma-Aldrich). The pantothenic acid was added ("EBSS + B₅") or not ("EBSS - B₅") depending on experimental context. Overnight cultures in Mueller-Hinton broth were normalized to an OD₆₀₀ of 0.05 and were used to inoculate wells in 96-well plates (10 μL in 150 μL). Growth was examined in the synthetic EBSS-based medium after 24 h using an Omega Microplate Reader with an atmospheric control unit module (BMG Labtech) at 42 °C in a modified atmosphere composed of 10% CO₂ and 5% O₂.

ACKNOWLEDGMENTS. We thank Dr. Bruce Pearson for helpful comments and advice. S.K.S. is a Wellcome Trust Fellow. This work was carried out with support from the Wellcome Trust and the Biotechnology and Biological Sciences Research Council.

- Woolhouse ME, Taylor LH, Haydon DT (2001) Population biology of multihost pathogens. *Science* 292(5519):1109–1112.
- Giraud A, et al. (2001) Costs and benefits of high mutation rates: Adaptive evolution of bacteria in the mouse gut. *Science* 291(5513):2606–2608.
- Moxon R, Bayliss C, Hood D (2006) Bacterial contingency loci: The role of simple sequence DNA repeats in bacterial adaptation. *Annu Rev Genet* 40:307–333.
- Kim JS, et al. (2012) Passage of *Campylobacter jejuni* through the chicken reservoir or mice promotes phase variation in contingency genes Cj0045 and Cj0170 that strongly associates with colonization and disease in a mouse model. *Microbiology* 158(Pt 5): 1304–1316.
- Killiny N, Almeida RPP (2011) Gene regulation mediates host speci. city of a bacterial pathogen. *Environ Microbiol Rep* 3(6):791–797.
- Gottesman S (1984) Bacterial regulation: Global regulatory networks. *Annu Rev Genet* 18:415–441.
- McCarthy ND, et al. (2007) Host-associated genetic import in *Campylobacter jejuni*. *Emerg Infect Dis* 13(2):267–272.
- Sheppard SK, et al. (2011) Niche segregation and genetic structure of *Campylobacter jejuni* populations from wild and agricultural host species. *Mol Ecol* 20(16):3484–3490.
- Griekspoor P, et al. (2013) Marked host specificity and lack of phylogeographic population structure of *Campylobacter jejuni* in wild birds. *Mol Ecol* 22(5):1463–1472.
- Didelot X, Falush D (2007) Inference of bacterial microevolution using multilocus sequence data. *Genetics* 175(3):1251–1266.
- Parkhill J, et al. (2000) The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* 403(6770):665–668.
- Griggs DJ, et al. (2009) Beta-lactamase-mediated beta-lactam resistance in *Campylobacter* species: Prevalence of Cj0299 (bla OXA-61) and evidence for a novel beta-lactamase in *C. jejuni*. *Antimicrob Agents Chemother* 53(8):3357–3364.
- Pearson BM, et al. (2003) Comparative genome analysis of *Campylobacter jejuni* using whole genome DNA microarrays. *FEBS Lett* 554(1–2):224–230.
- McDonald P (2011) *Animal Nutrition* (Prentice Hall/Pearson, Harlow, England), 7th Ed.
- Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447(7145): 661–678.
- Wellcome Trust Case Control Consortium; Craddock N, et al. (2010) Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 464(7289):713–720.
- Balouz S, Francis AR, Lan R, Tanaka MM (2010) Conditions for the evolution of gene clusters in bacterial genomes. *PLoS Comput Biol* 6(2):e1000672.
- Hacker J, Kaper JB (2000) Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* 54:641–679.
- Falkow S (1988) Molecular Koch's postulates applied to microbial pathogenicity. *Rev Infect Dis* 10(Suppl 2):S274–S276.
- Beres SB, et al. (2004) Genome-wide molecular dissection of serotype M3 group A Streptococcus strains causing two epidemics of invasive infections. *Proc Natl Acad Sci USA* 101(32):11833–11838.
- Jolley KA, Maiden MC (2010) BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11:595.
- Sheppard SK, et al. (2013) Progressive genome-wide introgression in agricultural *Campylobacter coli*. *Mol Ecol* 22(4):1051–1064.
- Pearson BM, et al. (2007) The complete genome sequence of *Campylobacter jejuni* strain 81116 (NCTC11828). *J Bacteriol* 189(22):8402–8403.
- Fouts DE, et al. (2005) Major structural differences and novel potential virulence mechanisms from the genomes of multiple campylobacter species. *PLoS Biol* 3(1):e15.
- Parkhill J, et al. (2001) Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* 413(6858):848–852.
- Lefebvre T, Bitar PD, Suzuki H, Stanhope MJ (2010) Evolutionary dynamics of complete *Campylobacter* pan-genomes and the bacterial species concept. *Genome Biol Evol* 2:646–655.
- Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18(5):821–829.
- Cabello H, et al. (1997) Bacterial colonization of distal airways in healthy subjects and chronic lung disease: A bronchoscopic study. *Eur Respir J* 10(5):1137–1144.
- Gundogdu O, et al. (2007) Re-annotation and re-analysis of the *Campylobacter jejuni* NCTC11168 genome sequence. *BMC Genomics* 8:162.
- Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.
- Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 9(4): 299–306.
- Falush D, Bowden R (2006) Genome-wide association mapping in bacteria? *Trends Microbiol* 14(8):353–355.
- Sims GE, Jun SR, Wu GA, Kim SH (2009) Alignment-free genome comparison with feature frequency profiles (FFP) and optimal resolutions. *Proc Natl Acad Sci USA* 106(8):2677–2682.
- Sims GE, Kim SH (2011) Whole-genome phylogeny of *Escherichia coli*/Shigella group by feature frequency profiles (FFPs). *Proc Natl Acad Sci USA* 108(20):8329–8334.
- Marchini J, Cardon LR, Phillips MS, Donnelly P (2004) The effects of human population structure on large genetic association studies. *Nat Genet* 36(5):512–517.
- Martins EP, Garland T (1991) Phylogenetic analyses of the correlated evolution of continuous characters: A simulation study. *Evolution* 45(3):534–557.
- Garland T, Jr., Bennett AF, Rezende EL (2005) Phylogenetic approaches in comparative physiology. *J Exp Biol* 208(Pt 16):3015–3035.
- Garland T, Dickerman AW, Janis CM, Jones JA (1993) Phylogenetic analysis of covariance by computer simulation. *Syst Biol* 42:265–292.
- Rutherford K, et al. (2000) Artemis: Sequence visualization and annotation. *Bioinformatics* 16(10):944–945.
- Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J (2009) DNAPlotter: Circular and linear interactive genome visualization. *Bioinformatics* 25(1):119–120.