Point mutations in the major outer membrane protein drive hypervirulence of a rapidly expanding clone of *Campylobacter jejuni*

Zuowei Wu[a], Balamurugan Periaswamy[b,c], Orhan Sahin[a,d], Michael Yaeger[e], Paul Plummer[a,d], Weiwei Zhai[c], Zhangqi Shen[a], Lei Dai[a], Swaine L. Chen[b,c,1], and Qijing Zhang[a,1]

[a]Department of Veterinary Microbiology and Preventive Medicine, Iowa State University, Ames, IA 50011; [b]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119074; [c]Genome Institute of Singapore, Singapore 138672; [d]Department of Veterinary Diagnostic and Production Animal Medicine, Iowa State University, Ames, IA 50011; and [e]Department of Veterinary Pathology, Iowa State University, Ames, IA 50011

Edited by John J. Mekalanos, Harvard Medical School, Boston, MA, and approved July 26, 2016 (received for review April 12, 2016)

Infections due to clonal expansion of highly virulent bacterial strains are clear and present threats to human and animal health. Association of genetic changes with disease is now a routine, but identification of causative mutations that enable disease remains difficult. *Campylobacter jejuni* is an important zoonotic pathogen transmitted to humans mainly via the foodborne route. *C. jejuni* typically colonizes the gut, but a hypervirulent and rapidly expanding clone of *C. jejuni* recently emerged, which is able to translocate across the intestinal tract, causing systemic infection and abortion in pregnant animals. The genetic basis responsible for this hypervirulence is unknown. Here, we developed a strategy, termed "directed genome evolution," by using hybridization between abortifacient and nonabortifacient strains followed by selection in an animal disease model and whole-genome sequence analysis. This strategy successfully identified SNPs in *porA*, encoding the major outer membrane protein, are responsible for the hypervirulence. Defined mutagenesis verified that these mutations were both necessary and sufficient for causing abortion. Furthermore, sequence analysis identified *porA* as the gene with the top genome-wide signal of adaptive evolution using Fu's Fs, a population genetic metric for recent population size changes, which is consistent with the recent expansion of clone "sheep abortion." These results identify a key virulence factor in *Campylobacter* and a potential target for the control of this zoonotic pathogen. Furthermore, this study provides general, unbiased experimental and computational approaches that are broadly applicable for efficient elucidation of disease-causing mutations in bacterial pathogens.

clonal expansion | bacterial pathogenesis | pathogen evolution | population genetics | *Campylobacter*

Pathogens evolve relentlessly, often resulting in regional or global expansion of successful clones or strains. In many of these cases, pathogens causing outbreaks are hypothesized to have increased virulence, due to acquisition of new transmission, survival, or infection traits (1–7). Knowledge of the precipitating genetic and phenotypic change(s) responsible for virulence is necessary for guiding rational design of effective control measures. Whole genome sequencing has provided a powerful tool to identify such evolutionary genetic changes and has transformed the ways by which we understand bacterial virulence, pathogenesis, epidemiology, and evolution (8). However, the elucidation of exact mechanisms underlying the success of individual pathogenic clones remains difficult, especially for recently emergent and expanding clones with enhanced virulence. In these cases, we are often left only with correlated genetic markers and lack information on causative changes. This is especially true when the success of a pathogenic clone may only involve minor genetic changes, such as SNPs.

Campylobacter jejuni is a major foodborne pathogen and a leading cause of enteritis in humans, responsible for 400–500 million cases of diarrhea annually worldwide (9). In the United States alone, *Campylobacter* accounts for more than 800,000 cases of foodborne illnesses each year (10). As a zoonotic pathogen, *C. jejuni* is widely distributed in the gut microbiota of wild and domesticated animal species, such as cattle, sheep, and poultry (11, 12). Transmission of *C. jejuni* to humans is mainly via contaminated meat, milk, and water. Although *C. jejuni* is primarily a gut colonizer, some hypervirulent strains may be able to translocate across intestinal epithelium, producing bacteremia and systemic infections (13). In addition to causing foodborne illnesses, *C. jejuni* is also a primary etiological agent for ruminant abortion (14). Recently, we reported the emergence of an antibiotic-resistant and hypervirulent clone of *C. jejuni* in the United States (15). It is named clone "sheep abortion" (SA) and is responsible for the majority (>90%) of ovine abortion cases in the United States (15). The clonality of SA strains was established by multiple molecular typing methods, including pulsed-field gel electrophoresis (PFGE) and multilocus sequence typing (MLST) (ST-8) (15). Notably, *C. jejuni* clone SA is zoonotic and has been implicated in a number of outbreaks and sporadic cases of foodborne illnesses in humans (16). Clone SA infection is characterized by its extraordinary ability to translocate across the intestine, induce systemic infection, invade the uteroplacental unit with high titer, and cause abortion in pregnant animals compared with other strains of *C. jejuni* (15, 17).

Significance

Identification of specific genetic changes responsible for pathogen emergence and evolution is critical for disease prevention and control. By using a powerful and high throughput approach (directed genome evolution), we identified the specific point mutations in the major outer membrane protein that drive the hypervirulence of an emergent *Campylobacter jejuni* clone, which causes abortion in ruminants and foodborne disease outbreaks in humans. This finding reveals a critical virulence factor and a potential target for the control of *Campylobacter*, an important zoonotic pathogen affecting both animal and human health. The experimental and computational approaches developed in this study are generally applicable to other bacterial organisms for identifying specific virulence factors responsible for a disease phenotype.

Author contributions: Z.W. and Q.Z. designed research; Z.W., O.S., M.Y., P.P., Z.S., and L.D. performed research; Z.W., B.P., O.S., W.Z., and S.L.C. analyzed data; and Z.W., B.P., S.L.C., and Q.Z. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: Reads data were deposited in the Sequence Read Archive (SRA) under accession SRP067965.

[1]To whom correspondence may be addressed. Email: zhang123@iastate.edu or slchen@gis.a-star.edu.sg.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1605869113/-/DCSupplemental.

Epidemiological analysis of historical clone SA isolates suggested that its emergence in the United States is likely to be a recent event (15), but the genetic basis for its emergence and enhanced virulence is unknown. To facilitate understanding the pathogenesis, we sequenced a representative clone SA isolate (IA3902), which revealed that the genome of IA3902 is highly syntenic to that of *C. jejuni* strain NCTC11168 (18). Importantly, *C. jejuni* NCTC11168 was shown to be nonabortifacient in pregnant animals (17), providing a closely related control strain for elucidating the hypervirulence of clone SA. Comparative genomics revealed numerous genetic differences between the genomes of IA3902 and NCTC11168, including 57 unique genes and >8,000 SNPs as well as small (<10 bp) insertions and deletions (indels). Transcriptomic and proteomic analyses identified multiple differences in the expression of genes involved in several classical virulence-related pathways: iron acquisition, capsule biosynthesis, energy metabolism, and motility (18). However, the causative genetic changes driving clone SA's success in causing disease remained undetermined.

In this study, we integrated genomics with experimental approaches to reach beyond correlative analyses to identify the exact genetic changes responsible for hypervirulence. We first sequenced a panel of *C. jejuni* strains, which established the emergence and rapid expansion of *C. jejuni* clone SA across the United States. Next we developed the "directed genome evolution" strategy, which takes advantage of transformation between two genetically similar but phenotypically different (abortifacient and nonabortifacient) strains, positive screening in an animal model of the hybrids that gain virulence, and whole genome sequence analysis to pinpoint the locus and mutations responsible for the disease phenotype. This strategy effectively identified mutations in a single causative gene (*porA*; encoding the major outer membrane protein) (MOMP) driving the abortion phenotype of clone SA. Subsequent experiments verified that point mutations in a single, surface-exposed loop (loop 4) of the MOMP were both necessary and sufficient for causing abortion in the guinea pig model. Armed with this definitive genetic evidence, we evaluated several computational tests for adaptive evolution using the genome sequencing data, finding that Fu's Fs test singled out *porA* as the gene with the strongest genome-wide signal for adaptive evolution. This finding represents a rare example of arguing that genome sequence analysis can, in an unbiased fashion, produce clear hypotheses for causative genetic changes.

Results and Discussion

Comparative Genomics and Verification of Clone SA Virulence. In total, 99 *C. jejuni* isolates were sequenced, including 72 clone SA isolates (ST-8) and 27 nonclone SA isolates (non–ST-8). They were selected from our collections over the last two decades to represent historical and contemporary isolates of clone SA as well as sporadic nonclone SA abortion isolates in the United States and Great Britain (*SI Appendix*, Table S1 and Fig. S1). A whole genome phylogeny definitively confirmed the previous MLST data (15, 16) that clone SA has undergone a recent, monophyletic clonal expansion into different hosts (cattle, goat, chicken, and human). This finding was supported by the tree topology (Fig. 1*A*), reduced recombination subsequent to emergence (*SI Appendix*, Fig. S2), and a reduced haplotype diversity (median h = 0 for clone SA versus 0.794 for nonclone SA isolates). In contrast, the nonclone SA abortion isolates are paraphyletic and genetically indistinguishable from nonabortion reference strains (Fig. 1*A* and *SI Appendix*, Fig. S3). Molecular clock analysis led to an estimate of its emergence in the mid 1970s (Fig. 1*B*). This finding is comparable with the time detected for the species shift from *Campylobacter fetus* to *C. jejuni* in the etiology of sheep abortion (15, 16), which could be due to the deployment of *C. fetus* vaccines. Additionally, we performed animal infection experiments using clone SA and nonclone SA isolates, which showed IA3902 and D7324 (clone SA isolates) were highly abortifacient, whereas NCTC11168 and VDL2401 (nonclone SA isolates) did not induce any abortion in the inoculated guinea pigs (Fig. 1*C*). The differences in abortion rates were statistically significant ($P < 0.05$). These findings are consistent with the results from previous studies (17) and demonstrated the distinct ability of clone SA in abortion induction. The enhanced virulence of clone SA in inducing abortion has likely facilitated its transmission in sheep flocks as aborted materials (fetus, placenta, uterine discharge, etc.) serve as an important source of infection for healthy ewes. Thus, the hypervirulence is not only a marker for clone SA, but may also be linked to its emergence as the predominant cause of sheep abortion in the United States.

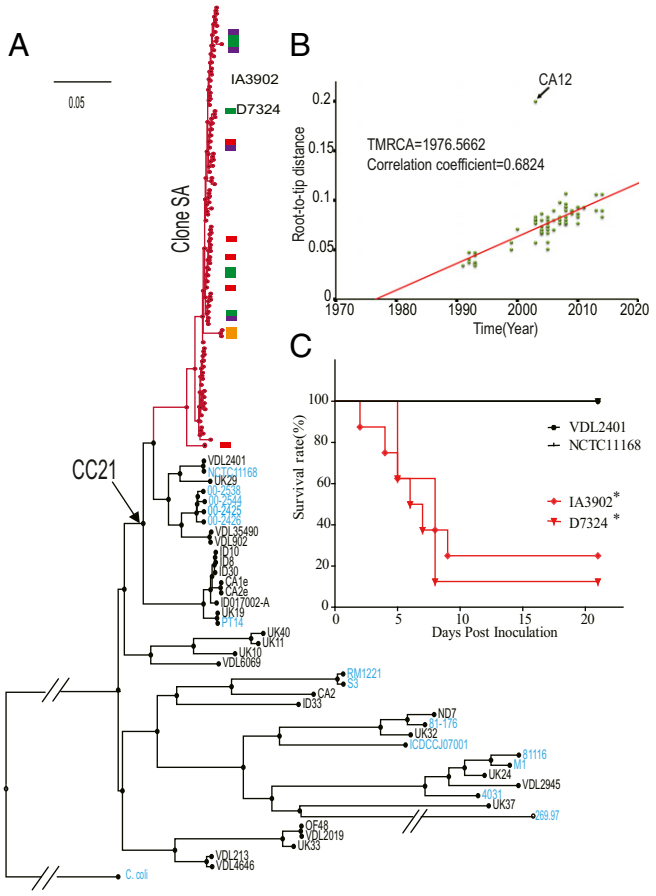

Fig. 1. Population genetics analysis and virulence properties of *C. jejuni* abortion isolates. (*A*) Maximum-likelihood phylogenetic tree based on whole genome SNP differences of 114 *C. jejuni* isolates and rooted to *Campylobacter coli*. Taxon names of clone SA isolates are not shown, but their branches are highlighted in red; and isolates from nonsheep sources are indicated by filled box of red (bovine abortion), purple (caprine abortion), orange (chicken meat), and green (human gastroenteritis). Nonclone SA references are indicated by blue. (Scale bar, number of nucleotide substitutions per site.) Root and strain 269.97 lengths are not to scale. (*B*) Regression analysis between isolation dates (*x* axis) and root-to-tip distance (*y* axis). The analysis was conducted based on 73 clone SA isolates. The point where the line intersects with the *x* axis gives the inferred date when the most recent common ancestor of *C. jejuni* clone SA emerged. The lone outlier CA12 is indicated by an arrow. (*C*) Virulence assessment of clone SA isolates (IA3902 and D7324) and nonclone SA isolates (VDL2401 and NCTC11168) in a pregnant guinea pig model. Survival curves (nonaborted pregnant/total pregnant) were compared by the log-rank (Mantel–Cox) test. Asterisk indicates a statistically significant difference from the NCTC11168-inoculated group ($P < 0.05$).

The monophyletic emergence of clone SA and its ability to induce abortion is thought to have been due to the ultimate acquisition of genetic changes enabling its virulence. Such changes could involve either small SNP/indels or larger events such as gene gain/loss due to horizontal gene transfer or chromosomal

MICROBIOLOGY

rearrangements. However, our initial effort focusing on examining gene contents by comparing clone SA with nonclone SA isolates did not identify plausible candidate genes that were unique to clone SA and might explain the hypervirulence (*SI Appendix*, Fig. S4). Thus, the causative loci responsible for the hypervirulence of clone SA remained unknown.

Design of a Directed Genome Evolution Strategy for Identifying Causative Mutations. We therefore devised an experimental strategy to identify the genetic changes causing abortion. This strategy involved generation of hybrids by natural transformation between two genomically similar, but phenotypically different *C. jejuni* strains, positive screening in an animal model of the hybrids (transformants) that gained virulence, and whole genome sequence analysis of the hybrids from infected tissues (Fig. 2*A*). We termed this strategy directed genome evolution, as it mimics genome-wide evolution guided by positive selection. The transformation was performed between *C. jejuni* NCTC11168 and a clone SA isolate IA3902, which were nonabortifacient and highly abortifacient, respectively, as demonstrated in a guinea pig model (Fig. 1*C*) (17). Purified genomic DNA of strain IA3902 was used as the donor DNA, whereas *C. jejuni* NCTC11168, which has a genome highly syntenic to that of IA3902 (18), was used as the recipient strain. Complete genome sequences are available for both strains, but 25 chromosomal genes and an extra plasmid are found to be specific for IA3902; in addition, 8,696 SNPs and indels were identified between these two strains (18, 19). Thus, individual analyses of the contribution of these differences to virulence would be time and cost prohibitive, which demanded for an effective and high throughput strategy for screening the candidate loci.

Nonabortifacient NCTC11168 was grown through serial passages in culture media in the continuous presence of purified IA3902 (abortifacient) genomic DNA. Transformation efficiency was monitored by measuring the frequency of tetracycline-resistant colonies. The recipient NCTC11168 is tetracycline susceptible, whereas IA3902 is resistant due to the possession of a chromosomal *tet(O)* gene (18). The frequency of tetracycline-resistant transformants plateaued after four passages (*SI Appendix*, Fig. S5), and this library of transformants was used for inoculation of guinea pigs.

Via oral gavage, three groups ($n = 8$ per group) of guinea pigs were inoculated with NCTC11168, IA3902, and the library of NCTC11168 transformants (hereafter NCTC11168-tr), respectively. Abortion in this model is associated with the ability to translocate across the intestinal barrier and induce infection of the uterus and placenta (17). The nonabortifacient NCTC11168 colonized the intestinal tract, but did not induce abortion, histological signs of inflammation in uterus and placenta, or infection of the uteroplacental unit as confirmed by bacterial culture. In contrast, animals infected with IA3902 or NCTC11168-tr yielded high and comparable ($P > 0.05$) abortion rates (Fig. 2*B*), with high bacterial loads (*SI Appendix*, Fig. S6*A*) and inflammation in uteroplacental units. This result clearly indicated that NCTC11168-tr gained the ability to induce abortion in pregnant guinea pigs. This finding was significant, as it suggested that the virulence-contributing genetic factor was successfully transferred from IA3902 to NCTC11168 via natural transformation.

Bacteria were isolated from the uterus and placenta of NCTC11168-tr–infected animals that had undergone abortion. We selected 12 individual isolates randomly picked from three aborted animals (4 per animal) as well as a pooled sample consisting of all remaining colonies from the uterus and placenta of all three animals. These isolates were from infected uterus and placenta and should have gained the virulence for abortion induction. This positive screening step in the guinea pig model ensured that those tranformants that did not gain virulence were blocked entering the uteroplacental unit. The positive selection was amazingly effective as plating with Tet-containing media demonstrated that the input transformant library (~10^9 cfu) contained a number of Tet-resistant colonies (~10^5 cfu), whereas

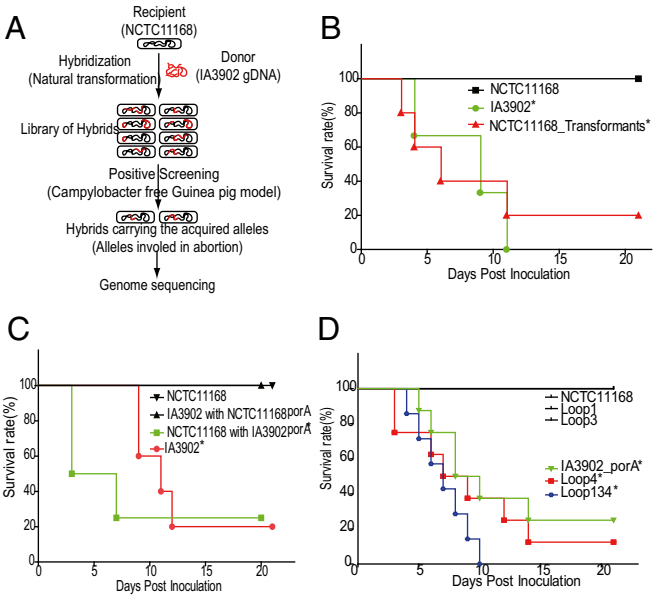

Fig. 2. Identification of *porA* mutations responsible for the hypervirulence of *C. jejuni* clone SA. (*A*) Schematic depiction of the directed genome evolution strategy applied in this study. (*B*) Gain-of-function screening (positive selection) of NCTC11168 transformants that acquired the ability to cause abortion in pregnant guinea pigs. Three groups of animals were orally challenged with IA3902, NCTC11168, and NCTC11168 transformants, respectively. (*C*) Defining the role of *porA* in virulence by allele exchanges between IA3902 and NCTC11168. Pregnant guinea pig groups were challenged with NCTC11168, IA3902, NCTC11168 with IA3902 *porA*, and IA3902 with NCTC11168 *porA*, respectively. (*D*) Identification of the specific SNPs of *porA* responsible for virulence. Groups of pregnant guinea pigs were challenged with NCTC11168, NCTC11168 with loop 1 of IA3902 *porA* (loop 1); NCTC11168 with loop 3 of IA3902 *porA* (loop 3), NCTC11168 with loop 4 of IA3902 *porA* (loop 4), NCTC11168 with loops 1, 3, and 4 of IA3902 *porA* (loop 134), and NCTC11168 with the entire IA3902 *porA* (IA3902_porA), respectively. The predicted external loops of the *porA*-encoded protein are shown in Fig. 3*C*. In *B*–*D*, the asterisk indicates a statistically significant difference from the NCTC11168-inoculated group ($P < 0.05$).

the pooled isolates from the infected uterus and placenta did not contain any Tet-resistant colonies. This finding indicates that the transformants that acquired *tet(O)* did not reach the uterus and placenta, consistent with the fact that *tet(O)* does not contribute to abortion induction, and also suggesting that *tet(O)* is not linked to a critical virulence factor required for abortion induction.

PFGE analysis demonstrated that the pool and all 12 individual isolates from the infected uterus and placenta were indeed NCTC11168 derived, as their PFGE patterns were identical to NCTC11168 (*SI Appendix*, Fig. S7). This PFGE result also suggested that no large chromosomal rearrangements occurred in the transformants. The individual isolates and the pool were subject to whole genome sequence analysis, which did not reveal transfer of IA3902-specific genes to the selected isolates. Then we focused on analysis of the transferred small mutations (SNPs and indels) using a conservative strategy (*SI Appendix*, Fig. S8), in which only reads supporting differences between the two genomes that could be unambiguously identified were used. In all samples, as expected, most of the positions carried the original NCTC11168 (nonabortion) allele. Among the 12 individual isolates, a total of 47 loci representing 279 SNPs and 15 indels (SNP/indels from either genic/intergenic regions) were transferred in at least one isolate. In particular, 12 loci were unanimously present in all 12 isolates (Fig. 3*A*). However, the mutations in 5/12 loci (*mreB*, *cheA*, *cj0431*, *cj0455c*, and *cj0807*) were also present in the parental NCTC11168 as resequenced in this study (Fig. 3*A*). Interestingly, in the pooled sample, only 7 of these 12 loci had high allele frequencies (>99%) for the IA3902 (abortion) allele,

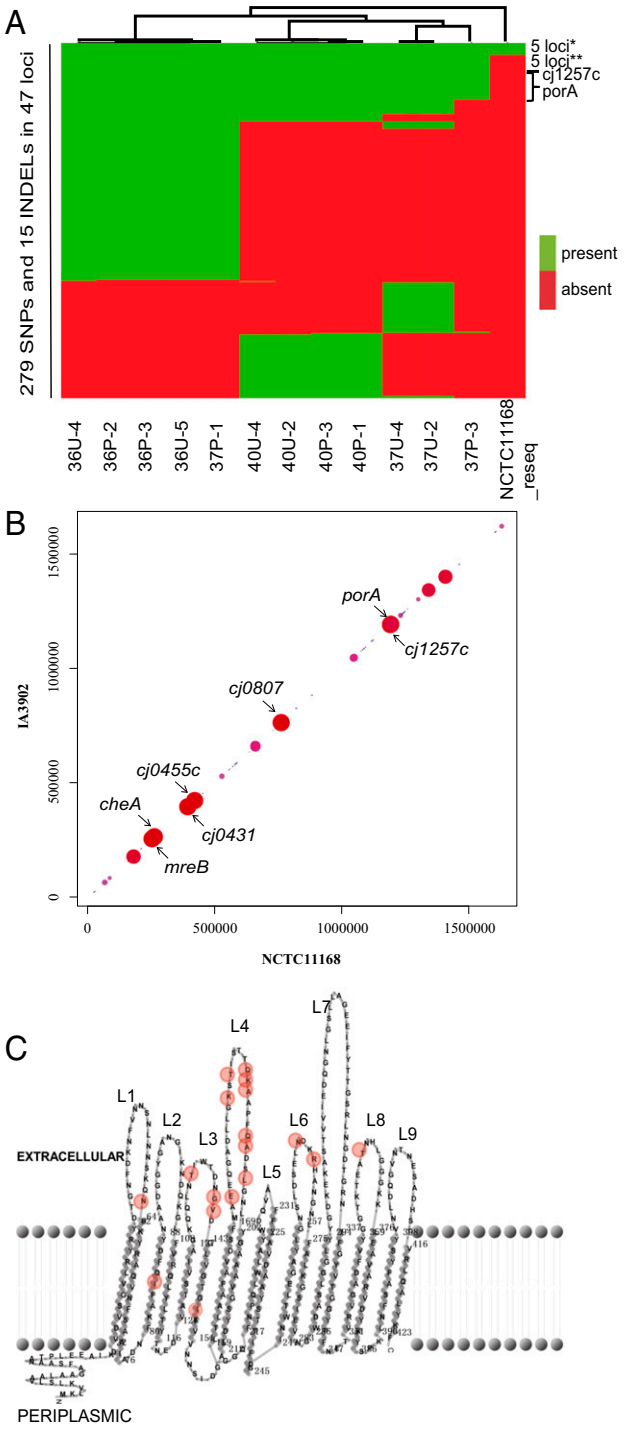

Fig. 3. Identification of small mutations (SNPs/indels) among the NCTC11168 transformants by Illumina sequencing 12 independent colonies and a pooled sample of colonies obtained from uterus and placenta of three aborted guinea pigs. (*A*) Heat map of 279 SNPs and 15 indels obtained from 12 independent colonies. The NCTC11168 recipient strain was resequenced in this study as a control. Each row represents a small mutation (SNP or indel) and colors indicate absence/presence (red/green) of the same mutation in a given sample (column labels). Mutations in 12 loci are unanimously present in all 12 colonies and the loci are *(*mreB*, *cheA*, *cj0431*, *cj0455c*, and *cj0807*), **(*cj0046*, *cj1470c*, *cj0184c*, and intergenic sites 1190424 and 1338383), *cj1257c*, and *porA*. (*B*) Plot of small mutations from the pooled sample (982 SNPs and 24 indels) using NCTC11168 as reference (*x* axis) and IA3902 (*y* axis). Mutational changes were collinear on the *x*–*y* axis with size of the dots proportional to allele frequency (red, high frequency; pink, intermediate frequency; blue, <1%), alleles with >99% frequency are indicated by arrows, and the related loci are shown. (*C*) Predicted secondary structures of MOMP (encoded by *porA*) of *C. jejuni* IA3902 by PRED-TMBB (41) and visualized by TMRPres2D (42). All of the amino acid level differences relative to *C. jejuni* NCTC11168 are highlighted by red circles.

including *porA*, *cj1257c*, and the 5 loci described above (Fig. 3*B*). The mutations in the remaining 5 loci did not have high frequencies in the pooled sample, although they were present in all 12 individual isolates (Fig. 3*A*). Therefore, only 2 loci, which contained two genes: *cj1257c* and *porA*, had high allele frequencies in both the individual colonies and the pooled sample, and did not have the corresponding mutations in the parental NCTC11168. *cj1257c* differs between IA3902 and NCTC11168 by a synonymous SNP, whereas *porA* carries 16 synonymous and 25 nonsynonymous SNPs that lead to 18 aa substitutions (*SI Appendix*, Table S2). Therefore, our linkage analysis produced *porA* as the single most likely candidate gene for abortion. *porA (cjsa_1198)* and *tet(O) (cjsa_0193)* are distantly separated (by ~990 kb) in the chromosome of IA3902 and are unlikely cointegrated into the recipient during transformation, which further explains why *tet(O)* was absent in the transformants isolated from the uteroplacental units of aborted animals.

Specific Mutations in *porA* Are Required for *C. jejuni* Systemic Infection and Abortion. We further validated the findings from the directed genome evolution experiment by using defined mutagenesis and animal studies. First, directed allelic exchange was used to swap the chromosomal allele of *porA* in IA3902 and NCTC11168. IA3902 with the *porA* from NCTC11168 completely lost the ability to induce abortion, whereas NCTC11168 with the *porA* of IA3902 fully gained virulence in pregnant guinea pigs (Fig. 2*C*). In fact, the abortion rates between IA3902 and NCTC11168 with the *porA* of IA3902 were comparable and were not significantly different ($P > 0.05$). This result showed that the IA3902 *porA* allele was both necessary and sufficient for causing abortion in the context of existing genes of NCTC11168 (Fig. 2*C*).

porA encodes the MOMP in *C. jejuni* and MOMP has been shown to function as a porin and adhesion (20–22). MOMP is the most abundant membrane protein in *C. jejuni* with 18 β-strands and nine external loops (23). The 18 aa differences between the *porA* alleles of NCTC11168 and IA3902 clustered preferentially on several of the nine surface-exposed loops (Fig. 3*C*). We hypothesized that some of these mutations were required for abortion induction. To test the hypothesis, site-specific mutagenesis was used to introduce specific mutations into the loops of *porA* in NCTC11168. These mutant constructs were evaluated in the guinea pig model, which demonstrated that among the sets of mutations in loops 1, 3, and 4, only loop 4 mutations (containing 9 aa substitutions) were required and sufficient to confer an abortion phenotype when introduced into NCTC11168 (Fig. 2*D*). Again, in these experiments abortion was strictly correlated with *C. jejuni* infection in the uterus and placenta (*SI Appendix*, Fig. S6*B*) and histopathological signs of inflammation. The mutations in loop 6 and loop 8 were not tested because they have low frequencies among either the transformants or the field clone SA isolates. These findings indicate that allelic variations in loop 4 are critical for the hypervirulence of *C. jejuni* clone SA.

***porA* Drives the Recent Expansion of Clone SA.** The *porA* alleles among clone SA isolates were closely related to the IA3902 allele (*SI Appendix*, Fig. S10*A*). Strikingly, two nonclone SA strains (VDL902 and VDL35490), which were isolated from abortion cases, carried *porA* alleles very similar to clone SA strains. In addition, four clone SA isolates carried *porA* alleles divergent from the majority of the clone SA strains. Three of them (FDA17848, FDAN337, and FDAN342) were isolated from nonabortion cases; but the other one, CA12, was an exception, as it was from a bovine abortion case with two ~37-kb phage inserts in the genome compared with IA3902 and other clone SA strains. Although CA12 is

MICROBIOLOGY

categorized as a clone SA isolate based on its core genome, it is quite divergent from other clone SA isolates (Fig. 1*B*). Additionally, detailed analysis of the *porA* allele indicated *porA* of CA12 was quite different from that of clone SA, but was closely related to that of ICDCCJ07001 (a nonclone SA strain) (*SI Appendix*, Fig. S10), suggesting that recombination might have changed the *porA* allelic profile of CA12. These differences in the *porA* allele and whole genome phylogeny are likely due to recombination, which occurs commonly in *Campylobacter* (24, 25), and suggest that CA12 is an outlier of the clone SA clade.

Genome sequencing is widely used for studying evolution of bacterial pathogens (8, 26). The finding that allelic variation in loop 4 of *porA* is necessary and sufficient for conferring abortion to NCTC11168 presents a unique dataset in the sense of a positive control for evaluating sequence analysis methods that can directly provide mechanistic insights into disease. The epidemiology of clone SA suggests a recent selective sweep (driven by positive selection) acting on the *porA* gene (Fig. 4*A*). Several tests are designed to detect signals of recent selective sweeps or positive selection from sequence data (27–30). However, neither selective sweep nor codon-based positive selection tests found a signal at the *porA* locus that would stand out from all other genes in the genome. Thus, these methods were unable to find a signal associated with abortion for *porA* (*SI Appendix*, *SI Results*).

Because we sampled strains from different time periods and geographic locations, which might violate demographic assumptions of some population genetics analyses (31), we focused our analysis solely on clone SA isolates from known abortion cases. We reasoned that this would enable analysis on the relevant population after the selective sweep and population expansion and also remove the confounding effect of recombination at the *porA* locus. As noted above, among the 59 clone SA isolates known from abortion cases, one strain, CA12, had a divergent *porA*, likely due to recombination (*SI Appendix*, Fig. S10). Therefore, we performed analysis on datasets including and excluding CA12. Strikingly, in the nonrecombined dataset (excluding CA12), Fu's Fs test predicted *porA* as the top locus driving selection/population expansion (*SI Appendix*, Fig. S11*B*; see *SI Appendix*, *SI Results* for full details of analysis of all datasets). A typical signature of a recent selective sweep is an enrichment of recent, rare mutations that result in an excess of unique haplotypes at the selected locus (30, 32, 33). Indeed, we saw precisely this theoretical pattern at the *porA* locus of clone SA, centered on loop 4, explaining its outstanding signal using Fu's Fs test (Fig. 4 *B*–*D*). Inclusion of CA12 in the analysis would falsely increase the overall diversity of a clonally expanding population without an apparent increase in haplotype frequency. As Fu's Fs test relies on haplotype frequency, given the diversity of the population (30), this would result in a false negative signal for *porA*, justifying the exclusion of CA12 from the analysis.

Our finding of *porA* as the top hit using Fu's Fs provides a rare example, suggesting that genomic analysis may be capable of directly identifying disease-causing loci in silico. This success relied on outstanding metadata regarding abortion, clear epidemiology indicating population expansion, and an experimentally verified dataset on *porA* provided by the directed genome evolution strategy. We were able to manually examine our dataset for recombination, due to the knowledge that mutations in *porA* were responsible for abortion, underlying the need for better methods to automatically remove such confounding factors in the prediction of unknown loci. The ability of Fu's Fs for identifying the genetic basis for clone SA's success matches expectations gained from theory and simulation (30, 34, 35), indicating that further developments in characterizing recombination, population growth models, and sampling strategies could have broader impact on the value of other pathogen sequencing projects.

Concluding Remarks. Emergence of pathogenic variants is often driven by their increased virulence or fitness, possibly associated with the gain of novel genes or mutations (1). Identification of the genetic basis for altered virulence phenotypes is essential for developing better strategies to recognize, predict, and control disease emergence and epidemics (36). In recent years, development of next-generation sequencing technologies has greatly facilitated our understanding of pathogen evolution as well as genotype–phenotype relationships. In many cases, investigation focused on large regions of genetic variation, such as pathogenicity islands, recombination events, and mobile genetic elements that introduce novel genes to emergent and epidemic strains (1). Recent examples include acquisition of the streptococcal superantigen (SSA) and transposable elements encoding multidrug resistance genes driving the expansion of scarlet fever-associated *Streptococcus pyogenes* emm12 lineages in Hong Kong (5) and horizontal transfer of a mobile genetic element-encoded gene *sasX* driving epidemic waves of methicillin-resistant *Staphylococcus aureus* (MRSA) infection (4). Whereas SNPs are the most common sources of genomic variations within bacteria and also involved in the change of pathogen's virulence and host tropism (37–40), such mutations are rarely reported to drive clonal expansion of bacterial pathogens. It is likely that the sheer number of SNPs existing among strains of the same bacterial species, coupled with the paucity of methods for effective screening of causative mutations, has led to this bias toward large gene and mobile element differences in existing genomics studies.

In this study, we have discovered that specific amino acid substitutions in loop 4 of MOMP encoded by *porA* are responsible

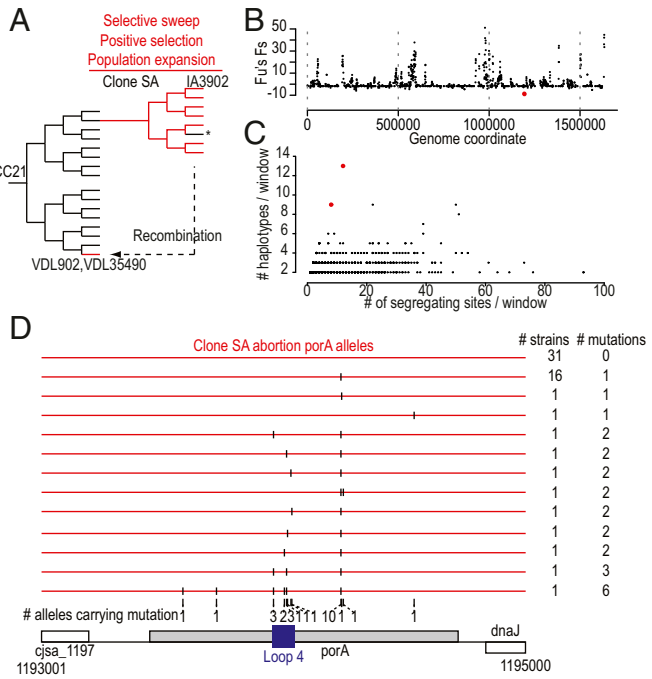

Fig. 4. Evolutionary analysis of *porA*. (*A*) Working model for clone SA evolution. Clone SA (red) has undergone clonal expansion recently in clonal complex (CC) 21, driven by a selective advantage conferred by *porA*. Ongoing recombination exerts both loss of highly abortifacient potential in clone SA (black star) and gain of the same in nonclone SA strains (VDL902 and VDL35490). (*B* and *C*) Genome-wide sliding window analysis (window size, 1 kb; step size, 500 bp). (*B*) Plot of Fu's Fs statistic (*y* axis) and the genomic coordinate of the center of the window (*x* axis). (*C*) Plot of the number of haplotypes (*y* axis) versus the number of segregating sites (*x* axis, clipped at 100). Windows containing *porA* are indicated with larger red dots (*B* and *C*). (*D*) Mutation map of the *porA* locus. Each allele (haplotype) of a 2-kb window surrounding *porA* is depicted (red line). The number of strains and the number of mutations differentiating that allele from the consensus is indicated (*Right*). Black vertical lines indicate the location of a mutation. Allele count per mutation per position is shown (*Bottom*). Those with one allele carrying the mutation are singletons. The genomic organization of *porA* loop 4 region (dark blue) and neighboring genes are depicted (*Bottom*).

for the hypervirulence of *C. jejuni* clone SA. How these minor changes lead to a major change in virulence is intriguing, but it is probably due to alteration of a key function of MOMP, which has been identified as a porin and adhesion (20–22). Additionally, glycan modification of MOMP was previously shown to be involved in intestinal colonization (20). However, these previously known functions of MOMP may not fully explain its key role in causing abortion and it is highly possible that MOMP has a yet-to-be-discovered role in mediating pathogen–host interaction and enabling systemic infection. This possibility remains to be determined in future studies.

Work presented in this study represents documentation of the key role of MOMP in systemic infection and abortion induction by *C. jejuni*. This finding is significant as it provides a potential target for controlling the spread of *C. jejuni* clone SA, which is a zoonotic pathogen causing abortion in ruminants and foodborne illnesses in humans (15, 16). Additionally, we developed a directed genome evolution strategy for effective screening of causative mutations, which may be applied to understand the virulence mechanisms of other pathogenic organisms that are amenable to the same strategy. Compared with traditional loss-of-function, gain-of-function, or random mutagenesis screens, this strategy has the primary advantage of being efficient and adaptable. By starting with phenotypically different strains, the causative genetic changes responsible for a phenotype can be readily identified using the strategy. Furthermore, our data present a unique case study for evaluating and refining methods used for genomic analyses. Finally, findings from this study provide direction for investigating pathogen–host interactions during infection by *C. jejuni*.

Materials and Methods

Bacterial isolates used in this study, whole genome sequencing, phylogenetic and temporal analyses, gene content analysis, construction of NCTC11168 transformants, animal experiments using pregnant guinea pigs, genome sequencing, and comparative genomics of *C. jejuni* transformants, allelic exchange between IA3902 and NCTC1168, and whole genome positive selection analysis are described in detail in *SI Appendix, SI Materials and Methods*. All animal experiments were approved by the Institutional Animal Care and Use Committee (IACUC) at Iowa State University.

ACKNOWLEDGMENTS. The authors thank Nada Pavlovic for technical support and Jun Lin (University of Tennessee) for discussions. This work was supported by Agriculture and Food Research Initiative's Animal Health Competitive Grant 2013-67015-20368 from the USDA National Institute of Food and Agriculture (to Q.Z., M.Y., and P.P.); the National Research Foundation, Singapore (Grant NRF-RF2010-10 to S.L.C.); the National Medical Research Council, Ministry of Health, Singapore Grant NMRC/CIRG/1357/2013 (to S.L.C. and B.P.); and the Genome Institute of Singapore/Agency for Science, Technology, and Research.

1. Morens DM, Fauci AS (2013) Emerging infectious diseases: Threats to human health and global stability. *PLoS Pathog* 9(7):e1003467.
2. Rohde H, et al.; E. coli O104:H4 Genome Analysis Crowd-Sourcing Consortium (2011) Open-source genomic analysis of Shiga-toxin-producing E. coli O104:H4. *N Engl J Med* 365(8):718–724.
3. Petty NK, et al. (2014) Global dissemination of a multidrug resistant *Escherichia coli* clone. *Proc Natl Acad Sci USA* 111(15):5694–5699.
4. Li M, et al. (2012) MRSA epidemic linked to a quickly spreading colonization and virulence determinant. *Nat Med* 18(5):816–819.
5. Davies MR, et al. (2015) Emergence of scarlet fever Streptococcus pyogenes emm12 clones in Hong Kong is associated with toxin acquisition and multidrug resistance. *Nat Genet* 47(1):84–87.
6. He M, et al. (2013) Emergence and global spread of epidemic healthcare-associated Clostridium difficile. *Nat Genet* 45(1):109–113.
7. Nicolas-Chanoine MH, et al. (2008) Intercontinental emergence of Escherichia coli clone O25:H4-ST131 producing CTX-M-15. *J Antimicrob Chemother* 61(2):273–281.
8. Wilson DJ (2012) Insights from genomics into bacterial pathogen populations. *PLoS Pathog* 8(9):e1002874.
9. Ruiz-Palacios GM (2007) The health burden of Campylobacter infection and the impact of antimicrobial resistance: Playing chicken. *Clin Infect Dis* 44(5):701–703.
10. Scallan E, et al. (2011) Foodborne illness acquired in the United States: Major pathogens. *Emerg Infect Dis* 17(1):7–15.
11. Rotariu O, et al. (2009) Spatiotemporal homogeneity of Campylobacter subtypes from cattle and sheep across northeastern and southwestern Scotland. *Appl Environ Microbiol* 75(19):6275–6281.
12. Hermans D, et al. (2012) Poultry as a host for the zoonotic pathogen Campylobacter jejuni. *Vector Borne Zoonotic Dis* 12(2):89–98.
13. Kaakoush NO, Castaño-Rodríguez N, Mitchell HM, Man SM (2015) Global epidemiology of Campylobacter infection. *Clin Microbiol Rev* 28(3):687–720.
14. Mannering SA, West DM, Fenwick SG, Marchant RM, O'Connell K (2006) Pulsed-field gel electrophoresis of Campylobacter jejuni sheep abortion isolates. *Vet Microbiol* 115(1-3):237–242.
15. Sahin O, et al. (2008) Emergence of a tetracycline-resistant Campylobacter jejuni clone associated with outbreaks of ovine abortion in the United States. *J Clin Microbiol* 46(5):1663–1671.
16. Sahin O, et al. (2012) Molecular evidence for zoonotic transmission of an emergent, highly pathogenic Campylobacter jejuni clone in the United States. *J Clin Microbiol* 50(3):680–687.
17. Burrough ER, Sahin O, Plummer PJ, Zhang Q, Yaeger MJ (2009) Pathogenicity of an emergent, ovine abortifacient Campylobacter jejuni clone orally inoculated into pregnant guinea pigs. *Am J Vet Res* 70(10):1269–1276.
18. Wu Z, et al. (2013) Multi-omics approaches to deciphering a hypervirulent strain of Campylobacter jejuni. *Genome Biol Evol* 5(11):2217–2230.
19. Parkhill J, et al. (2000) The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences. *Nature* 403(6770):665–668.
20. Mahdavi J, et al. (2014) A novel O-linked glycan modulates Campylobacter jejuni major outer membrane protein-mediated adhesion to human histo-blood group antigens and chicken colonization. *Open Biol* 4:130202.
21. Dé E, et al. (2000) MOMP (major outer membrane protein) of Campylobacter jejuni; a versatile pore-forming protein. *FEBS Lett* 469(1):93–97.
22. Moser I, Schroeder W, Salnikow J (1997) Campylobacter jejuni major outer membrane protein and a 59-kDa protein are involved in binding to fibronectin and INT 407 cell membranes. *FEMS Microbiol Lett* 157(2):233–238.
23. Zhang Q, Meitzler JC, Huang S, Morishita T (2000) Sequence polymorphism, predicted secondary structures, and surface-exposed conformational epitopes of Campylobacter major outer membrane protein. *Infect Immun* 68(10):5679–5689.
24. Wilson DJ, et al. (2009) Rapid evolution and the importance of recombination to the gastroenteric pathogen Campylobacter jejuni. *Mol Biol Evol* 26(2):385–397.
25. Sheppard SK, Maiden MC (2015) The evolution of Campylobacter jejuni and Campylobacter coli. *Cold Spring Harb Perspect Biol* 7(8):a018119.
26. Bentley SD, Parkhill J (2015) Genomic perspectives on the evolution and spread of bacterial pathogens. *Proc Biol Sci* 282(1821):20150488.
27. Alachiotis N, Stamatakis A, Pavlidis P (2012) OmegaPlus: A scalable tool for rapid detection of selective sweeps in whole-genome datasets. *Bioinformatics* 28(17): 2274–2275.
28. Pavlidis P, Živkovic D, Stamatakis A, Alachiotis N (2013) SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. *Mol Biol Evol* 30(9): 2224–2234.
29. Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics* 133(3): 693–709.
30. Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147(2):915–925.
31. Nielsen R (2001) Statistical tests of selective neutrality in the age of genomics. *Heredity (Edinb)* 86(Pt 6):641–647.
32. Messer PW, Neher RA (2012) Estimating the strength of selective sweeps from deep population diversity data. *Genetics* 191(2):593–605.
33. Oleksyk TK, Smith MW, O'Brien SJ (2009) Genome-wide scans for footprints of natural selection. *Philos Trans R Soc Lond B Biol Sci* 365(1537):185–205.
34. Ramírez-Soriano A, Ramos-Onsins SE, Rozas J, Calafell F, Navarro A (2008) Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics* 179(1):555–567.
35. Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Mol Biol Evol* 19(12):2092–2100.
36. Nasser W, et al. (2014) Evolutionary pathway to increased virulence and epidemic group A Streptococcus disease derived from 3,615 genome sequences. *Proc Natl Acad Sci USA* 111(17):E1768–E1776.
37. Viana D, et al. (2015) A single natural nucleotide mutation alters bacterial pathogen host tropism. *Nat Genet* 47(4):361–366.
38. Brault AC, et al. (2007) A single positively selected West Nile viral mutation confers increased virogenesis in American crows. *Nat Genet* 39(9):1162–1166.
39. Sanson M, et al. (2015) A naturally occurring single amino acid replacement in multiple gene regulator of group A Streptococcus significantly increases virulence. *Am J Pathol* 185(2):462–471.
40. Carroll RK, et al. (2011) Naturally occurring single amino acid replacements in a regulatory protein alter streptococcal gene expression and virulence in mice. *J Clin Invest* 121(5):1956–1968.
41. Bagos PG, Liakopoulos TD, Spyropoulos IC, Hamodrakas SJ (2004) PRED-TMBB: A web server for predicting the topology of beta-barrel outer membrane proteins. *Nucleic Acids Res* 32(Web Server Issue):W400–W404.
42. Spyropoulos IC, Liakopoulos TD, Bagos PG, Hamodrakas SJ (2004) TMRPres2D: High quality visual representation of transmembrane protein models. *Bioinformatics* 20(17):3258–3260.