A Link Between Virulence and Ecological Abundance in Natural Populations of *Staphylococcus aureus*

Nicholas P. J. Day,[1]* Catrin E. Moore,[1]
Mark C. Enright,[2]† Anthony R. Berendt,[1] John Maynard Smith,[3]
Michael F. Murphy,[4] Sharon J. Peacock,[5]
Brian G. Spratt[2]‡, Edward J. Feil[2]†

Staphylococcus aureus is a major cause of severe infection in humans and yet is carried without symptoms by a large proportion of the population. We used multilocus sequence typing to characterize isolates of *S. aureus* recovered from asymptomatic nasal carriage and from episodes of severe disease within a defined population. We identified a number of frequently carried genotypes that were disproportionately common as causes of disease, even taking into account their relative abundance among carriage isolates. The existence of these ecologically abundant hypervirulent clones suggests that factors promoting the ecological fitness of this important pathogen also increase its virulence.

Staphylococcus aureus is one of the most important bacterial pathogens of humans. The continuing burden of community- and hospital-acquired *S. aureus* disease, including serious endovascular, wound, bone, and joint infections (*1*, *2*), is a major public health concern. This concern is heightened by the increased prevalence of antibiotic-resistant strains such as methicillin-resistant *S. aureus* (MRSA) and glycopeptide-insensitive *S. aureus* (GISA) (*3*). Although up to 30% of the population of the United Kingdom carry *S. aureus* in their nostrils without symptoms, the annual reported incidence of bacteremia is less than 0.02% (*4*). Here, we address the question of whether all *S. aureus* are equally virulent, and infection is purely opportunistic, or whether invasive disease is primarily caused by a subset of particularly virulent genotypes unrepresentative of the carriage population as a whole.

For several pathogens that are carried asymptomatically, it has been shown that certain clones responsible for cases of invasive disease may also be abundant in the carriage population (*5*, *6*). However, in the absence of a representative population framework, it is not possible to deduce whether these clones are atypically virulent, or whether the apparent association between specific clones and disease simply reflects a higher rate of dissemination within the carriage population. To place isolates recovered from invasive *S. aureus* disease within the broader context of the species as a whole, we compared 61 bacterial isolates from patients with serious community-acquired invasive disease with 179 isolates recovered from the nostrils of healthy individuals living in the same community. We also studied isolates from 94 contemporaneous cases of hospital-acquired disease occurring in hospitals serving the study population, representing a different and clinically important epidemiological setting. All isolates were collected within Oxfordshire between 1997 and 1998 during a prospective case-control study to define host and bacterial factors associated with endemic invasive *S. aureus* disease (*7*).

We used multilocus sequence typing (MLST) to compare the isolates (*8*). Alleles at seven unlinked housekeeping loci are identified by sequencing ~450–base pair internal fragments of the genes (*9*, *10*), and the sequence type (ST) of an isolate is defined by the alleles at the seven loci. There is an average of 22.3 alleles per locus; hence, MLST could potentially resolve >1 billion STs. A clone is defined as a set of isolates identical at all seven loci.

The 334 isolates belonged to 187 STs; 26 STs were represented by more than one isolate (*11*). The carriage isolates were significantly more diverse than the disease isolates (*12*). Forty-eight percent (29/61) of community-acquired disease isolates belonged to just five STs, a degree of clonality consistent with the findings of a previous large study of disease isolates conducted using multilocus enzyme electrophoresis (*13*), whereas the five most common carriage STs account for just 14% (25/179) of the carriage isolate population. Thus, in the community, isolates causing disease are not drawn randomly from the carriage population.

Most of the isolates fall into clusters of closely related STs, or clonal complexes, and it is likely that isolates in each of these clusters have descended from a single ancestral genotype. To explore this possibility, we used an algorithm that first defines clonal complexes as groups in which each isolate is identical to at least one other isolate at five or more of the seven loci (*14*) (Fig. 1). In each of the 12 major clonal complexes, we have identified what we believe to be the "ancestral genotype," assigned as the ST differing from the highest number of other STs within the complex at only one locus. The ancestral genotype also corresponded to the largest clone in the clonal complex in 11 out of 12 cases, providing independent support for these assignments (*15*). After identifying putative ancestral genotypes, "single-locus variants" (SLVs) were identified. These are assumed to be direct descendents of ancestral genotypes, having undergone changes at a single locus, either by point mutation or recombination, but having remained identical to the ancestral genotype at the other six loci (*16*, *17*). A comparison of the frequencies of variant alleles within SLVs with their predicted ancestral counterparts provides further strong support for the ancestral assignments (*18*). In summary, a clonal complex is a set of genotypes derived in the recent past from a common ancestor. Clonal ancestors, and direct descendents of these ancestors, can be identified with some confidence within each clonal complex.

With two minor exceptions, all clonal complexes contained isolates recovered from community-acquired disease, nosocomial disease, and asymptomatic carriage. Thus, invasiveness is not a property of a few rare genotypes with unusually high virulence, and previous observations that particular pathogenic clones are commonly carried (*5*, *6*) can be extended to the bacterial population as a whole. Clones that are most virulent within the community have

[1]Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK. [2]Wellcome Trust Centre for the Epidemiology of Infectious Disease, University of Oxford, Oxford OX1 3FY. [3]School of Biological Sciences, University of Sussex, Falmer, Brighton, BN1 9QG, UK. [4]National Blood Service, John Radcliffe Hospital, Oxford OX3 9DU, UK. [5]Nuffield Department of Clinical Laboratory Sciences, University of Oxford, John Radcliffe Hospital, Oxford, OX3 9DU, UK.

*To whom correspondence should be addressed. E-mail: nick.day@ndm.ox.ac.uk
†Present address: Department of Biology and Biochemistry, University of Bath, Bath, BA2 7AY, UK.
‡Present address: Department of Infectious Disease Epidemiology, Imperial College School of Medicine, St. Mary's Campus, Norfolk Place, London, W2 1PG, UK.

also become common causes of disease within hospitals, in many cases having acquired resistance to multiple drugs in response to the hospital environment. Both methicillin-susceptible and MRSA isolates that cause hospital-acquired infections typically have ancestral genotypes, and there is also evidence for a similar trend in GISA isolates (*19*).

However, the crucial observation (Table 1) is that a significantly higher proportion of the disease isolates had an ancestral genotype [114/155 (74%), compared with 39/179 (22%) in the carriage sample, odds ratio for disease 10 (95% confidence intervals or CIs 5.5 to 18)]. Thus, even after taking into account their ecological abundance, isolates with ancestral genotypes (ancestral clones) remain disproportionately associated with disease. Furthermore, as ancestral clones have diversified by point mutation and recombination to form clonal complexes, there appears to have been an associated loss of virulence. Such an effect is more likely to be the result of recombination than point mutation, as recombination may also change loci neighboring the MLST genes. Evidence of an association between recombination and loss of virulence is apparent in the data; SLVs recovered from asymptomatic carriage were more likely to have arisen by recombination than those recovered from invasive disease ($P < 0.001$) (*18*). It is interesting that this loss of virulence was much more closely associated with recombination at

Table 1. Distribution of disease-causing and nasal carriage isolates within clonal complexes (see Fig. 1). A 4 by 2 Fisher's exact test comparing the distribution within the ancestral clones, single-locus variants, double-locus variants, and satellite strains of community-acquired disease with the distribution in nasal carriage isolates is highly significant ($P < 0.0001$). However, there is no statistical difference between isolates recovered from community- and hospital-acquired disease ($P = 0.86$).

Position within clonal complex	Disease isolates			Odds ratio (95% CI) for disease, relative to ancestral clones*
	Nasal carriage isolates [*n* (%)]	Community-acquired [*n* (%)]	Hospital-acquired [*n* (%)]	
Ancestral clone	39 (22)	45 (74)	69 (73)	1.00
Single-locus variant	68 (38)	9 (15)	18 (19)	0.11 (0.05 to 0.29)
Double-locus variant	31 (17)	4 (6)	3 (3)	0.11 (0.03 to 0.38)
Satellite†	16 (9)	0 (0)	1 (1)	0.00
Singleton (not within a clonal complex)	25 (14)	3 (5)	3 (3)	0.10 (0.03 to 0.41)

*This analysis includes only community-acquired disease and nasal carriage isolates, as these are epidemiologically directly comparable; hospital-acquired disease isolates are excluded. †A satellite is any isolate that is a member of a clonal complex, but is not a member of an ancestral clone or a single or double-locus variant (these exclude the strains within the three minor clonal complexes shown in Fig. 1 in which it is not possible to assign an ancestral clone; these are included in "singletons").

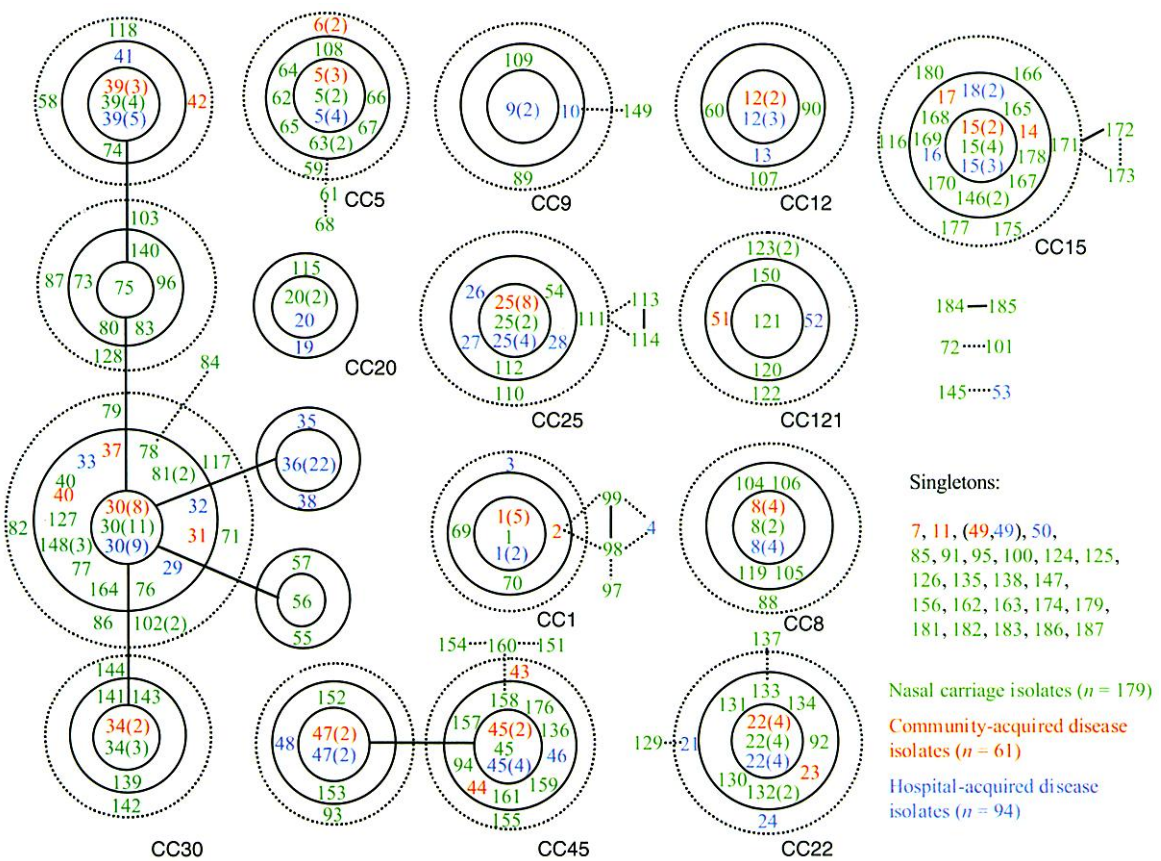

Fig. 1. Diagram of clonal complexes. Each number represents an MLST sequence type (ST). Where an ST is represented by multiple isolates, the number of isolates with that ST are shown in parentheses. Green numbers denote nasal carriage isolates, red numbers, community-acquired invasive disease isolates, and blue numbers, hospital-acquired disease isolates. No inferences are made concerning the relations between clonal complexes. The central circle of each clonal complex contains the ancestral clone of each clonal complex. Single-locus variants (SLVs) of an ancestral clone lie within the next (solid line) concentric circle, and double-locus variants within the outer (dotted line) circle. A solid straight line between two STs denotes a single-locus difference between them, a dashed straight line, a double-locus difference. Three pairs of related STs where the ancestral genotype cannot be predicted are also shown. Singletons are isolates possessing STs that differ from those of all other genotypes at >2 loci. In two of the clonal complexes, some SLVs were assigned as secondary ancestral clones because they differed by a single locus from at least two other genotypes that had not already been assigned as SLVs. Secondary ancestral clones and associated clonal variants were treated as primary ancestral clones in the analysis and were assigned in rank order according to the number of SLVs they define (for further details see the BURST readme file at www.mlst.net/BURST/BURSTREADME.htm). The clonal complexes are named according to the ST of the primary ancestral genotype, but with the prefix "CC" (for clonal complex). Isolates of ST1 (MSSA) and ST36 (EMRSA-16) are being sequenced at the Sanger Centre (*23*).

arcC, *tpi*, or *pta*, than with recombination at the other four loci, indicating that there may be virulence factors closely linked to these loci.

Why should the founding genotypes of clonal complexes be the most virulent, and why do clonal complexes exist at all? The presence of clones is often interpreted as reflecting low rates of recombination, but this is unlikely to be the case in *S. aureus* as we have shown that an allele is approximately eight times more likely to change by recombination than by point mutation (*18*). Alternatively, it is possible that ancestral genotypes carry a strong selective advantage, such that they spread sufficiently quickly to outrun the diversifying effects of recombination to become an observable clone (*20*, *21*).

If ancestral clones are both ecologically successful and disproportionately more likely to cause invasive disease, there may be a causal relation between fitness and virulence. In support of this, nasal carriage isolates with genotypes identical to those of invasive disease isolates were more likely to be recovered from both nostrils than those with genotypes unique to the carriage population (95% and 70%, respectively; $P = 0.002$). Similarly, carriage isolates within ancestral clones were more likely to colonize both nostrils than those not belonging to an ancestral clone (90% versus 71%, $P = 0.01$). These observations suggest that isolates corresponding to a virulent genotype and those belonging to ancestral clones are more successful colonizers than other genotypes.

Within clonal complexes the odds ratio for disease between ancestral clones and their putative descendants is 9.4 (95% CI, 5.0 to 17.6; $P < 0.0001$); that for colonization of both nostrils is 3.7 (1.1 to 12.5, $P = 0.02$). We speculate that greater virulence and greater propensity for colonization are pleiotropic effects of the same genetic change, and that, as ancestral clones diversify by point mutation and recombination to form clonal complexes, there is an associated loss of virulence and ecological fitness. As differential colonization of one or both nostrils is only an indirect indication of fitness, loss of fitness with clonal diversification cannot be demonstrated directly from the data.

As invasive disease is relatively very rare and unlikely to contribute to transmission to new hosts, enhanced virulence itself is unlikely to explain the ecological success of ancestral clones. An alternative explanation is that genetic factors promoting aggressive colonization also cause localized tissue damage, providing an increased likelihood of access to the blood stream and, hence, invasive disease. Such a link has been drawn between attachment factors (fimbriae) in *Escherichia coli* and urinary tract infection (*22*).

We conclude that clonal complexes arise in the natural staphylococcal population despite high rates of recombination, probably because the founding genotypes of these complexes carry a strong selective advantage. These founding clones are initially highly virulent; however, as the clones diversify (predominantly via recombination), the ability to cause invasive disease declines rapidly. This analysis highlights the importance when studying commonly carried bacterial pathogens of placing disease isolates in the context of the epidemiologically relevant carriage population. In the case of *S. aureus*, the results suggest that hypervirulent clones are abundant in the bacterial carriage population and that *S. aureus* is not solely an opportunistic pathogen.

References and Notes

1. T. G. Emori, R. P. Gaynes, *Clin. Microbiol. Rev.* **6**, 428 (1993).
2. J. P. Steinberg, C. C. Clark, B. O. Hackman, *Clin. Infect. Dis.* **23**, 255 (1996).
3. T. L. Smith *et al.*, *N. Engl. J. Med.* **340**, 493 (1999).
4. Public Health Laboratory Service Communicable Disease Surveillance Centre, *Communicable Dis. Rep.* **10**, 23 (2000).
5. J. M. Musser *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 225 (1990).
6. F. R. Cockerill *et al.*, *JAMA* **277**, 38 (1997).
7. Between July 1997 and July 1998 all cases of invasive *S. aureus* infection were identified prospectively through the microbiology laboratory serving the main hospitals in the Oxford area. Clinical details were recorded, and the bacterial isolate was stored. Invasive *S. aureus* disease was defined as the isolation of the organism from a normally sterile site in a patient with clinical signs and symptoms consistent with *S. aureus* infection; community-acquired disease was defined as admission to hospital with an illness consistent with invasive *S. aureus* disease, with isolation of *S. aureus* from a specimen taken from a normally sterile site within 24 hours of admission. Patients hospitalized for any reason within the preceding 6 months were excluded from this category.

   The 179 carriage isolates were recovered from nasal swabs taken from 547 healthy blood donors living in the catchment area of these hospitals and representative of the same population as those with invasive disease. Consent for nasal swabbing was sought from blood donors attending mobile blood donation clinics in 10 locations within this catchment area. The sites for these clinics were village halls, churches, and schools, none of which were usually used as a health-care facility. Two swabs were taken on the same occasion from each donor, one from each nostril.
8. M. C. Enright, N. P. Day, C. E. Davies, S. J. Peacock, B. G. Spratt, *J. Clin. Microbiol.* **38**, 1008 (2000).
9. M. C. Maiden *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3140 (1998).
10. B. G. Spratt, *Curr. Opin. Microbiol.* **2**, 312 (1999).
11. Sequence types for all the individual strains are available on www.mlst.net/ and in Web table 1, which is available in supplementary material on *Science* Online at www.sciencemag.org/cgi/content/full/292/5514/114/DC1
12. Forty-eight percent ($n = 161$) of all isolates had a unique ST, but carriage isolates were significantly more diverse [127 (71%) had unique STs] than those from community-acquired disease [13 (21%) had unique STs]. The probability of two isolates with identical STs being drawn at random from the carriage isolate population is 0.0056 (95% exact CI, 0.0045 to 0.0069), significantly lower than that of drawing two identical isolates from the community-acquired invasive disease population [0.049 (0.040 to 0.060)]. Probabilities were calculated as the proportion of all possible pairwise comparisons of isolates, where both isolates share the same ST. Exact binomial confidence limits were calculated for this proportion.
13. J. M. Musser, R. K. Selander, in *Molecular Biology of the* Staphylococci, R. Novick, R. A. Skurray, Eds. (VCH, New York, 1990), pp. 59–67.
14. This analysis was performed using the computer algorithm BURST, which is available on www.mlst.net/BURST/burst.htm
15. This BURST analysis was performed in two stages. Initially, the community-acquired disease and community nasal carriage groups of isolates were compared, as this is the primary comparison in the case-control study. Subsequently, the hospital-acquired disease isolates were added to the analysis. As this resulted in no significant changes to the clonal structure or to any of the central conclusions of this paper, we have, for the sake of brevity, only presented the latter analysis containing all three isolate groups (Fig. 1).
16. E. J. Feil, M. C. Maiden, M. Achtman, B. G. Spratt, *Mol. Biol. Evol.* **16**, 1496 (1999).
17. E. J. Feil, J. M. Smith, M. C. Enright, B. G. Spratt, *Genetics* **154**, 1439 (2000).
18. A total of 95 isolates differ from an ancestral genotype at only one of seven loci. These SLVs represented 87 different STs. Of these, 52 of the variant alleles differed from the allele in the ancestral clone at several nucleotide sites; these are almost certainly recombinants [see (*16*, *17*)]. The alleles in the remaining 35 SLVs involved only a single nucleotide change. A total of 21 of the variant alleles in these 35 SLVs were novel, that is, not present elsewhere in the data set, and we assigned these as having arisen by recent de novo point mutation. In contrast, all of the corresponding 35 ancestral alleles (present in the ancestral clones) were found elsewhere in unrelated lineages, suggesting that they predate the variant alleles in the SLVs ($P < 0.001$). This highly significant difference is strong evidence that the ancestral clones are indeed ancestral to the SLVs. The remaining 14 SLVs (involving single nucleotide changes resulting in a variant allele already present elsewhere in the data set) were assigned as recombinants. Thus, for the data set as a whole, we estimate that 21 SLVs have been generated by point mutation and 66 by recombination, suggesting that alleles change by recombination approximately 3 times as frequently as by point mutation [see (*16*, *17*) for a more detailed discussion of this method]. However, SLVs that have retained virulence are significantly more likely to have arisen by point mutation than SLVs from asymptomatic carriage (14/26 and 7/61, respectively; $P < 0.001$). If isolates recovered from invasive disease are removed, providing a more representative sample of the natural population, recombination is approximately eight times as frequent as point mutation. A table of ancestral genotypes and their SLVs, with details of the variant loci and alleles, is available as Web table 2 in supplementary material on *Science* Online at www.sciencemag.org/cgi/content/full/292/5514/114/DC1
19. M. C. Enright, unpublished data.
20. J. Maynard Smith, N. H. Smith, M. O'Rourke, B. G. Spratt, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4384 (1993).
21. J. Maynard Smith, E. J. Feil, N. H. Smith, *BioEssays* **22**, 1115 (2000).
22. J. G. Kusters, W. Gaastra, in *Fimbriae, Adhesion, Genetics, Biogenesis, and Vaccines*, P. Klemm, Ed. (CRC Press, Boca Raton, FL, 1994), p. 179.
23. More information on the sequencing of these two isolates can be obtained from www.sanger.ac.uk/Projects/S_aureus/.

24. This work was supported by The Wellcome Trust of Great Britain. We thank the staff of the National Blood Service in Oxford for their help with donor recruitment. We are grateful to R. Moxon, C. Newbold, N. White, and T. Peto for constructive comments on the manuscript and to M. Maiden for helpful discussions. The algorithm used to produce Fig. 1 was implemented with BURST, written by E.J.F. and M.-S. Chan.

11 October 2000; accepted 13 February 2001