At the smallest spatial scale, bacterial populations consist of a colony, a clonal entity produced by binary fission. The extent to which neighboring cells or colonies in the soil environment are genetically distinct has received relatively little attention. The few available studies report the coexistence of genetically different strains at very small scales. For example, up to 15 species were found in just 1 cm$^3$ of soil (59). Studies currently under way will also characterize phenotypic and behavioral diversity present in the same population.

*M. xanthus* is a gram-negative soil bacterium best known for its remarkable life history of social swarming, social predation, and multicellular fruiting body formation. Very little is known about genetic diversity within this species or how social strategies might vary among neighboring strains at small spatial scales. To investigate the small-scale population structure of *M. xanthus*, 78 clones were isolated from a patch of soil (16 by 16 cm) in Tübingen, Germany. Among these isolates, 21 genotypes could be distinguished from a concatemer of three gene fragments: *csgA* (developmental C signal), *fibA* (extracellular matrix-associated zinc metalloprotease), and *pilA* (the pilin subunit of type IV pilus). Accumulation curves showed that most of the diversity present at this scale was sampled. The *pilA* gene contains both conserved and highly variable regions, and two frequency-distribution tests provide evidence for balancing selection on this gene. The functional domains in the *csgA* gene were found to be conserved. Three instances of lateral gene transfer could be inferred from a comparison of individual gene phyllogenies, but no evidence was found for linkage equilibrium, supporting the view that *M. xanthus* evolution is largely clonal. This study shows that *M. xanthus* is surrounded by a variety of distinct conspecifics in its natural soil habitat at a spatial scale at which encounters among genotypes are likely.
the MLST method to a truly free-living prokaryote (i.e., not associated with a host), the first two being studies on thermophilic archaea (38a, 61). In MLST, several marker loci dispersed over the chromosome are sequenced to get a high-resolution representation of the genome. Relatively conserved housekeeping genes are usually chosen, but loci with greater variation can be used to detect diversity among more closely related strains (9). To provide an initial glimpse of genetic variation within \textit{M. xanthus}, five housekeeping genes and one nonessential gene were sequenced in 20 randomly chosen strains and three genes encoding cell surface proteins known to be important (csgA, pilA) or potentially important (fibA) in social interactions were sequenced in all isolates.

The \textit{csgA} gene codes for the cell surface-bound C-signal morphogen (29). Upon starvation-induced aggregation, it is transmitted by end-to-end contact between cells and induces aggregation by modifying the movement behavior of cells. The C signal is produced through positive feedback between cells and increases 100-fold over the course of development. High C-signal density at the final stage of fruiting body formation induces spore formation (22). The \textit{pilA} gene encodes the pilin subunit of type IV pili (28). Type IV pili are polar filaments that can be up to 10 \textmu m in length and are responsible for social gliding motility, presumably by attachment to a substrate and subsequent retraction, pulling the cell forward along its long axis (28). Extensive sequence variation in the \textit{pilA} gene has been reported in \textit{Neisseria} (1) and \textit{Pseudomonas} (54), genera with a \textit{pil} operon homologous to that of \textit{Myxococcus} (22). The \textit{fibA} gene codes for a putative zinc metalloprotease associated with the extracellular matrix surrounding cells (24). \textit{M. xanthus} produces an extracellular fibril matrix composed of approximately equal amounts of protein and carbohydrate (2). Fibril synthesis is stimulated by cell-cell contact and starvation and is essential for both social motility and fruiting body formation (24).

In contrast to non-sequence-based typing methods, MLST data can directly be interpreted in an evolutionary framework. Past recombination events can be investigated by testing for linkage disequilibrium (a nonrandom association among genes), and many sequence-based methods exist to analyze within-gene recombination (39). Whether recombination is an important factor in the evolution of \textit{M. xanthus}, or even whether it occurs at all, is unknown. Self-replicating plasmids are not retained by laboratory strains DK1622 and DZZ (19), but insertion of an \textit{Escherichia coli} plasmid into the \textit{M. xanthus} chromosome has been shown to mediate conjugative transfer between \textit{M. xanthus} cells (4). Phage attachment sites are present on the chromosome, and laboratory transductions are routinely performed (31). The importance of conjugation, phage transduction, or possible natural transformation in the life history of \textit{M. xanthus} has not been determined. Lateral gene transfer (localized sex) is very important in the evolution of many prokaryotes since it allows the direct acquisition of novel alleles or genes without having to evolve them de novo, thereby opening up new phenotypic abilities to a recipient strain. Since some of the genes sequenced in this study play important roles in the social interactions between cells, two population genetic tests were employed to examine whether natural selection might have acted to preserve (negative selection) or diversify (balancing selection) ancestral alleles.

**MATERIALS AND METHODS**

**Soil sampling and clone isolation.** Soil samples were collected in July 2003 in a wooded park lot in Tübingen, Germany (48°32'N, 9°3'E), from underneath dense shrubs, making it unlikely that the soil had been recently disturbed by human activity. Leaf litter was carefully removed, and 100 samples (labeled A0 to A99) were collected in 10 rows of 10 samples with sample edges separated by 1 cm in rows and columns (total surface area, 16 by 16 cm = 256 \text{cm}^2). Soil was collected by pushing sterile, open-end syringes (1 ml) into the ground, and the open end of each soil core was sealed with Parafilm. The diameter of the soil cores was 4.5 mm, and core lengths ranged from 5 to 20 mm.

The day after sampling, soil cores were crumbled and dispensed on selective agar (23) with sterile forceps. The upper and lower parts of the soil core were discarded to rule out possible contamination from the ground surface or the Parafilm, respectively. CTT (Casitone-Tris) medium (10 g of Casitone, 5 g of agar, 10 ml of 0.8 M MgSO\textsubscript{4}, 10 ml of 1 M Tris-HCl [pH 7.6], distilled H\textsubscript{2}O to 1,000 ml) was supplemented with vancomycin, neomycin, cycloheximide, and crystal violet (10-, 17.9-, 50-, and 10-g/ml final concentrations, respectively). These compounds do not target gram-negative bacteria and therefore imposed no selective bias on the isolation of different \textit{Myxococcus} strains. Plates were incubated at 32°C and 90% relative humidity and checked regularly under a dissecting microscope for the presence of fruiting bodies.

Fruiting bodies were picked from soil particles with sterile toothpicks. Toothpick tips were cut off and placed in 1.5-ml tubes containing 0.5 ml of distilled H\textsubscript{2}O. Samples were incubated at 30°C for 2 and sonicated with 2 for 10 s by using a tip sonicator to kill nonspores and to disperse spores (36). Spore suspensions were diluted into melted CTT soft agar (50°C) supplemented with antibiotics as described above, one colony (derived from a single spore) was randomly picked and transferred to a new selective plate. Finally, clones were grown in CTT liquid medium (at 32°C and 300 rpm) for DNA isolation and frozen storage (~80°C in 20% glycerol). Genomic DNA was isolated with an MBI Fermentas genomic DNA purification kit.

**PCR amplification and DNA sequencing.** Segments of the \textit{pilA}, \textit{csgA}, and \textit{fibA} genes were sequenced in all 78 clones. Five housekeeping genes, often utilized in MLST studies, were sequenced in a randomly chosen subset of 20 strains (listed in Table 1), i.e., GTP pyrophosphokinase (\textit{gppK}), ATP-binding subunit (\textit{cpxP}), isocitrate dehydrogenase (\textit{icd}), sigma factor 70 (\textit{rpoD}), and phosphoglucomutase (\textit{pgp}), as well as a nonessential DnaK homologue (HSP70 chaperone) (\textit{hspK}). In 11 of these clones, the 16S rRNA gene was sequenced to confirm species identity (listed in Table 1). Both strands were sequenced in all genes, except for the \textit{csgA} gene, where only one primer was used. Sequences of corresponding segments of the same 10 genes were retrieved from the genome sequence of the well-characterized laboratory strain DK1622 to serve as outgroups (BLAST search available at TIGR-CMR website [http://pathema.tigr.org/tigr-scripts/CPR/GenomePage.cgi?org=mgm]). Figure 1 shows the position of each gene on the DK1622 chromosome.

**PCR mixtures included 10 \mu l of 5X PCR buffer, 10 \mu l of 5X GC-Melt, 5 \mu l of Clontech Advantage GC Genomic Taq polymerase mix (Clontech Advantage GC genomic polymerase kit), 200 ng of genomic template DNA, 1 \mu l (10 pmol) of each primer, 5 \mu l of 2 mM deoxynucleoside triphosphates, and high-performance liquid chromatography grade H\textsubscript{2}O. Sequencing reactions were performed on 6.5 \mu l of 20-fold-diluted PCR product with 0.5 \mu l of AB 3.1 BigDye, 1 \mu l (5 pmol) of primer, and 2 \mu l of sequencing buffer. Sequencing reactions made use of PCR or internal primers. Primer sequences and PCR conditions are available upon request. Sequence trace files were evaluated with Seqman II software (DNASTAR, Madison, WI) and subsequently converted to FASTA format.

**Sampling effort.** To determine how much of the genetic diversity in the local population was sampled, accumulation curves were made with the program EstimateS version 6b1a, developed by R. K. Cobwell (http://www.berkley.edu/EstimateS) (8). An accumulation curve is the product of both the diversity of a population and the sampling effort. If enough samples are taken, the same genotypes will be sampled repeatedly and the accumulation curve reaches saturation, indicating that no more genotypes are likely to be present.

**Spatial analysis.** Because the nature of functional variation in bacteria cannot be readily inferred, even from considerable amounts of sequence data, we treated every genotype as an independent unit. A Student t test was employed to test whether distinct genotypes (based on the csgA-pilA-fibA concatenated) were distributed randomly in the spatial sampling grid. First, the overall frequency of each genotype occurring more than once in the grid was calculated. Second, for each isolated clone, the total number of neighbors and the number of identical neighbors were scored. Neighbors were defined as all clones surrounding a given clone (a maximum number of eight). For each clone, this yielded observed and
expected frequencies of identical neighbor clones, which were used to test for an excess of identical neighbor clones.

**Phylogenetic analysis.** DNA sequences were aligned with the ClustalW algorithm implemented in MEGA, version 3.0 (27) (www.megasoftware.net), except for the highly variable pilA gene fragment, for which the web-based protein alignment algorithm MUSCLE v6.0 (12) was used (www.drive5.com/muscle).

After trimming sequences to the maximum shared length (in reading frame), analyses were performed on individual gene trees, a concatemer of all nine genes sequenced in all clones, a concatemer of nine genes sequenced in a subset of clones (see above), and the same concatemer excluding additional genes were sequenced were the previous 11 plus A44, A45, A46, A47, A48, A49, A51, A53, and A58. was used to construct trees and networks. Bootstrap tests of phylogeny were performed with 1,000 replicates.

**MLST analysis.** Analyses of alleles (not sequences) were performed with applications made available on the MLST homepage (www.mlst.net). Allele assignments were made for each locus through the MLST database program NRDB. The 20 strains for which nine gene fragments were available were thus assigned a combination of alleleic types (numbers), known as a sequence types (STs). The eBURST algorithm (15) is designed to cluster STs together in so-called clonal complexes. A clonal complex emerges when a founder clone increases in frequency in the population, either because of a selective advantage or because of random drift. This clone will diversify over time and radiate into a number of offspring clades that differ at one of the sequenced loci. These single-locus variants are grouped around a strain assigned to be the founder clone of the clonal complex by the eBURST algorithm. Double-locus variants are linked to single-locus variants and thus differ from a founder clone. The START program (21), available through the MLST website, was used to perform the index-of-association (IA) test (53) to assess the level of linkage disequilibrium. This statistical test attempts to measure the extent of linkage equilibrium within a population by quantifying the amount of recombination among a set of sequences and detecting association between alleles at different loci. All STs are listed in Table S1 in the supplemental material.

**Tests of recombination.** The START program (21) was used to perform the maximum chi-squared test (52) and Sawyer’s runs test (47). Both tests are nucleotide substitution distribution methods that test for clustering of polymorphisms along an alignment (21, 39). Putative recombination breakpoints are identified, and permuted data sets are used to test for significance. Two phylogenetic sliding-window methods, implemented in the program TOPALi (33) (http://www.bioss.ac.uk/%7Eeainm/topali/), were used to predict possible recombination breakpoints.

**Tests of selection.** Ewens (13) derived a probability distribution for allele frequencies under the infinite-alleles model in a neutrally evolving population. Watterson (60) devised a test to see if an actual sample of allele frequencies deviates significantly from this distribution. The test statistic $F_{st}$ is the probability that two alleles chosen at random will be the same. A significantly small $F_{st}$ value indicates diversifying selection, whereas a significantly large $F_{st}$ value indicates purifying selection. The program Arlequin (48) (http://lgb.unige.ch/arlequin) was used to perform two versions of this test, the Ewens-Watterson test and the Ewens-Watterson-Slatkin test.

Applying the same logic as in the Ewens-Watterson test, the relative frequency of sites rather than alleles can be tested for accordance with a neutrally evolving population in equilibrium. The null model in population genetics considers
nucleotide polymorphism to be a product of effective population size \((N_e)\) and mutation rate \((\mu)\) only. Since the rate of introduction of new mutations in two randomly compared alleles is 2 \(\mu\), at neutral equilibrium nucleotide variation equals \(2N_e\mu\) in haploid organisms (this is called the mutation parameter \(M\)). The two measurements of DNA polymorphism nucleotide diversity \((\pi)\) and the number of segregating sites \((S)\) (or rather Watterson’s theta \([\theta]\); the number of segregating sites, \(S\), divided by Watterson’s constant, \(a_w\), to correct for sample size dependence) both lead to different estimates of \(M\) when the neutral theory is not satisfied (i.e., selection takes place). Segregating sites are equivalent to polymorphic sites, whereas mutations are equivalent to polymorphisms. This distinction is made because in some cases more than one polymorphism can be present at a particular polymorphic site. The programs MEGA 3.0 and DnaSP 4.0 (http://www.ub.es/dnasp) were used to conduct Tajima’s test.

### RESULTS

#### Isolation.
Between 1 and 2 weeks of incubation, bright yellow fruiting bodies \((\textit{xanthus} \text{ means yellow})\) emerged from 78 (out of 100) soil samples (Fig. 2). Fruiting bodies are in the size range of hundreds of micrometres and can be seen with the naked eye. Occasionally, swarming on the agar surface was observed. In these swarms, wave-like ripples of cells characteristic of the genus \textit{Myxococcus} \((49)\) could be observed before proper fruiting body formation.

The 16S rRNA sequences from 11 randomly chosen clones were identical (Table 1). These 11 clones were subsequently resolved into nine distinct genotypes on the basis of nine sequenced gene fragments. The strains in this study differed from the laboratory strain DK1622 by a single point mutation in \(>1,000\) bp of 16S rRNA, allowing us to assign our isolates to the species \textit{M. xanthus}.

#### Genetic diversity.
Our sequencing effort \((1,388 \text{ bp per clone with an additional 2,922 bp for the 20 randomly chosen clones})\) allowed us to distinguish 22 unique genotypes in total. Almost all genotypes were resolved with the \textit{csgA}, \textit{fibA}, and \textit{pilA} sequences alone, as the six additional genes sequenced in 20 randomly chosen clones yielded only one extra genotype beyond those resolved by the \textit{csgA}, \textit{fibA}, and \textit{pilA} sequences (clone A59 differed at the \textit{pgi} locus). Table 1 summarizes information on all of the genes sequenced in this study. The genes \textit{csgA}, \textit{fibA}, and \textit{pilA} appear twice: first with data on all 78 of the clones sequenced and then with data on the subset of 20 clones for which sequences of the six additional genes were determined as well. When comparing the number of \textit{csgA}, \textit{fibA}, and \textit{pilA} alleles found among all strains to that found in the subset of 20, it is clear that allele numbers are higher because of the larger sample size rather than increased diversity in these genes (see below). The \textit{csgA} gene yielded the most alleles \((n = 13)\) among the 78 clones, but only 7 \textit{csgA} alleles remained when just 20 clones were considered.

By far the highest number of polymorphisms was found in the \textit{pilA} gene: 135 polymorphisms plus six indels (insertions or deletions). The proportion of polymorphisms in the rest of the genes sequenced in the subset of 20 clones ranged from 0.7% to 2.4%, whereas \textit{pilA} was polymorphic at 35.5% of the sites among these clones. The two measurements of DNA polymorphism nucleotide diversity \((\pi)\) and the number of segregating (polymorphic) sites corrected for sample size \((\theta)\) are more formal descriptions of molecular diversity and are included for comparison with other studies.

#### Sampling effort and spatial analysis.
To assess the degree to which our collection of 78 clones is representative of the total diversity of the \textit{M. xanthus} population within the sampled soil patch, accumulation curves were made for the \textit{csgA}, \textit{fibA}, and \textit{pilA} genes and the \textit{csgA-fibA-pilA} concatemer (Fig. 3). It is clear that the curve slopes substantially approach zero well before the last sample taken, indicating that most of the common \textit{csgA}, \textit{fibA}, and \textit{pilA} diversity present in the total population is represented within our collection of 78 clones. The concatemer curve levels off less abruptly but nonetheless indicates that most of the common genotypes in the grid were sampled. A logarithmic regression curve fit predicts a total of 26 genotypes (rather than the 20 actually found) if 200 clones (rather than 78 clones) had been isolated and genotyped.

A Student \(t\) test was not quite supportive of a nonrandom (clustered) distribution of the different genotypes \((P = 0.0864, \text{ df } = 140)\). The two most numerous genotypes \((\text{each } n = 15)\) were not significantly clustered either \((P = 0.3363\) and 0.3774, respectively, by the Student \(t\) test).
Phylogenetic analysis. Figure 4 shows an unrooted NJ tree based on a concatemer of the csgA, fibA, and pilA genes for all 78 strains. Although it can give a misleading rooted appearance, the tree is depicted in phylogram format to enable a clear listing of all 78 isolates. The 21 genotypes resolved by this concatemer can be divided into six deeply branching groups that are exclusively defined by large genetic distances between the different pilA alleles. The pilA gene tree is therefore entirely congruent with the concatemer tree. Average genetic distances within the csgA and fibA gene phylogenies are 22- and 13-fold lower than the pilA phylogeny, respectively (Table 1). These differences in genetic resolution (branch length) make it difficult to directly draw conclusions about patterns of possible incongruence among the three gene trees in Fig. 5. However, three clear instances of incongruence appear to provide strong evidence for past recombination events among M. xanthus genomes.

First, in the pilA phylogeny (and therefore in the concatemer phylogeny), A98 clusters in group I (represented by strains A0 and A25 in Fig. 4 and 5). In the csgA phylogeny, however, strain A98 clusters with group VI (strains A66 and A75), which stands out as distantly related to all of the other groups in all three gene phylogenies. A98 is also more closely related to group VI strain A66 than to the group I strains in the fibA phylogeny. Second, strain A75 possesses one of the common fibA alleles and does not cluster with the other group VI strains in the fibA phylogeny. Third, laboratory strain DK1622, the ancestor of which was isolated decades ago in California, is closely related to group IV (represented by A12 in Fig. 5) in the pilA phylogeny but not in the fibA and csgA phylogenies (Fig. 5). These very apparent topology differences among gene trees can be taken as direct evidence for lateral gene transfer between strains. Because of this evidence for recombination, the concatemer tree depicted in Fig. 4 should not be used to infer phylogenetic relationships between strains.

Split decomposition analysis (20) was performed on all nine of the genes sequenced. Unlike traditional phylogenetic methods, split decomposition analysis does not impose a branching structure on the data set. It takes into consideration possible alternative connections between taxa that are, by definition, omitted from phylogenetic trees. This may result in a reticulated structure, with taxa connected by multiple edges (branches) and therefore internal nodes that do not represent ancestral genotypes. This statistical, not evolutionarily explicit, approach allows the extraction of conflicting phylogenetic signals that can be investigated in greater detail. The weight of each split is represented by its length so that only relatively square boxes represent instances in which both signals are equally strong. Recombination is often inferred when competing splits receive equal bootstrap support (e.g., see reference 46), but this interpretation must be made with extreme caution since homoplasy (similarity not caused by co-ancestry), sampling error (small number of sites under consideration), and systematic error (wrong model of sequence evolution) can also result in conflicting phylogenetic signals (20).

Split decomposition networks of the six genes sequenced in a subset of 20 strains displayed widely varying topologies that were usually not congruent (results not shown). The low number of polymorphisms in these genes, combined with the use of a distance-based method, might cause this lack of congruence rather than recombination. The clpX and relA phylogenies are bifurcating only and so are not indicative of recombination, whereas the icd, pgi, sglK, and rpoD networks do contain splits...
Laboratory strain DK1622 is a proper outgroup in the csgA, fibA, and pilA phylogenies (Fig. 5). However, when only the subset of 20 strains is considered, DK1622 forms an outgroup in the csgA, fibA, icd, pgi, and sglK trees (but not in the clpX, pilA, relA, and rpoD trees), suggesting that the majority of the strains are closely related and relatively far removed from DK1622 within M. xanthus. A split decomposition network of the concatemer sequence was constructed without the pilA gene, which would bias the network because of its high level of polymorphism (Fig. 6). The concatemer network makes clear the overall distant relationship of DK1622 with the cluster of 10 genotypes found in the 20 randomly sampled clones. However, if the clones highly divergent in the csgA, fibA, and pilA sequences (A98 and the four group VI strains) had been included in this subset, the picture would probably be less consistent. Only two splits are present in the concatemer network, one of which has equal weights but low bootstrap support. Different models of nucleotide substitution yielded very similar results in all trees and networks (not shown).

**MLST analysis.** Ten different STs could be distinguished among the 20 randomly chosen clones for which nine loci were sequenced (six alleles per locus, on average; Table 1). Visual inspection reveals that certain alleles are shared by clones that are otherwise very different. However, most of the alleles are unique, making it possible in only two cases (csgA and pgi) to group strains with the eBURST algorithm. For the other genes, too few alleles were shared between the other strains for them to be grouped together. Four point mutations separate A45 and A53 at the csgA locus, and one point mutation separates A59 from A15, A16, A31, A44, and A51 at the pgi locus. These groupings do not represent true clonal complexes, because more than two strains are needed to assign one of them as the most likely founder clone.

The combination of few shared alleles and low nucleotide divergence between strains (except for the pilA gene) may be indicative of past recombination. When a purely clonal organism evolves, mutations will accumulate at various loci over time. As a result, the number of alleles and the number of nucleotide polymorphisms at every locus should be positively correlated across strains (14). When recombination occurs frequently, this correlation disappears because alleles are exchanged between strains regardless of how much nucleotide divergence exists between them. Surprisingly, a negative correlation actually exists between the number of different alleles and nucleotide divergence in our data set. However, because only a few STs differ at a small number of loci, this result is likely due to stochasticity.

The I_A test (53) compares the observed variance in the distribution of allelic mismatches in all pairwise allelic profile comparisons to that expected in a freely recombining population. The I_A test is thus based on recombination between, rather than within, populations.

---

**FIG. 5.** NJ trees of the csgA (A), fibA (B), and pilA (C) gene fragments. One or more clones were selected as representatives of each major clade in the csgA-fibA-pilA concatemer (Fig. 4), and laboratory strain DK1622 was included for comparison. The bootstrap value (1,000 replicates) is given at each node. The corresponding roman numeral group designations used in Fig. 4 for all of the strains depicted are as follows: A0, A25, and A98, I; A5, II; A1, A2, A3, A4, A9, and A53, III; A12, IV; A17, V; A66 and A75, VI. Trees are not drawn to the same scale, and values in the upper left corner are genetic distances calculated with the Kimura two-parameter distance model.
than within genes. The $I_A$ statistic is calculated as $V_0/V_K - 1$, where $V_0$ is the observed variance and $V_K$ is the expected variance of $K$ (the number of loci at which two individuals differ). A value close to zero indicates linkage equilibrium (extensive recombination). The test was performed on the 10 STs based on the nine loci sequenced in 20 randomly chosen clones (Table 1). Notwithstanding the lack of correlation between nucleotide similarity and the number of shared alleles, the $I_A$ value of 2.164 is indicative of linkage disequilibrium (clonal evolution).

**Tests of recombination.** In addition to the test for linkage disequilibrium, several sequence-based tests of recombination were performed. It is important to use a combination of methods since the detection abilities of different tests can vary markedly for a given data set (39). The maximum chi-squared test was performed for every gene on all possible pairwise combinations (1,000 randomizations; significance test was performed for every gene on all possible pairwise combinations), and the test was performed on the 10 STs based on the nine loci sequenced in 20 randomly chosen clones (Table 1). Notwithstanding the lack of correlation between nucleotide similarity and the number of shared alleles, the $I_A$ value of 2.164 is indicative of linkage disequilibrium (clonal evolution).

**Polymorphisms in the csgA, fibA, and pilA gene fragments.** The protein alignment of the 13 csgA alleles corresponds to amino acid positions 15 to 206 of the p25 version of the DK1622 protein (29). Four nonsynonymous substitutions are present, but the amino acid sequences of the catalytic site are identical in all sequences. The sequence of the upstream coenzyme binding pocket is available in two strains and does not differ from that of DK1622.

The alignment of the six Tübingen fibA alleles corresponds to amino acid positions 123 to 284 in DK1622 (24). This is a region upstream from the putative active-site residues, and so nothing can be said about the functional significance of the eight amino acid changes present in the alignment.

A fragment length of 324 bp is used in all pilA analyses. Longer sequences were obtained in enough clones to allow comparison of one representative of each pilA genotype (nine total) over a longer region spanning amino acid positions 19 to 139. Knowledge of type IVa pilin structure is mainly derived from studies of Neisseria gonorrhoeae GC pilin and Pseudomonas aeruginosa PAK and K122-4 pilin (10). A comparison of the primary amino acid sequences and protein structures of these species allows us to make some inferences about the structure of M. xanthus pilin. All nine sequences are almost identical to DK1622 at residues 19 to 91, which cover an N-terminal α-helix that is conserved to allow tight packing of pilin subunits in the pilus hydrophobic core (10). Beyond position 91, the sequence is highly polymorphic, with especially group VI (Fig. 4 and 5) diverging from the other sequences. At position 92 (group IV and DK1622) or positions 92 and 93 (all others), there is a deletion relative to group VI. Another deletion is found at positions 128 and 129 in group VI.

This variable region is likely to be part of the globular head domain, which contains regions at the pilus surface that are likely to interact directly with extracellular matter, including other M. xanthus cells. One such region previously described in N. gonorrhoeae and P. aeruginosa (10) is the so-called D region, a loop between two cysteine residues connected by a disulfide bond. Two cysteines are present at positions 98 and 115 and are hypothesized to define a similar D region in M. xanthus. This putative D region is somewhat longer than the P. aeruginosa PAK and K122-4 pilin D regions (17 amino acids instead of 13). If this is truly a D region, it is located very close to the α-helix compared to other pilins (fewer than 10 residues, compared to 76 residues in PAK and K122-4 pilin) (10). Divergent group VI interestingly lacks both cysteines and thus this putative D region.

**Tests of selection.** A sufficiently large number of csgA, fibA, and pilA sequences was sampled from the population to provide confidence that our sampled allele frequencies approximate actual frequencies (Fig. 3). A nonbiased sample is important for testing the action of natural selection on genes. We used two frequency-distribution tests for selection on csgA, fibA, and pilA, which all encode surface-associated proteins.

Two versions of the Ewens-Watterson test were performed on these three genes to test for deviation from the null model of neutral evolution (Table 2). Since the $F$ statistic is the probability that two alleles chosen at random will be the same,
a lower-than-expected \( F \) value indicates that multiple alleles are present at high frequency and provides evidence for the operation of balancing selection (genetic polymorphism maintained by natural selection). Lower-than-expected \( F \) values are observed in the \( \text{fibA} \) and \( \text{pilA} \) genes, but only in \( \text{pilA} \) is this value significant (Watterson test, almost significant in the Slatkin test). A higher-than-expected \( F \) value, as found in the \( \text{csgA} \) gene, is indicative of purifying selection rather than balancing selection. The most common \( \text{csgA} \) allele was more abundant than expected (35 copies present where only 24 were expected) but did not yield a significantly low \( F \) value.

Approximately equal values for measurements of genetic variation based on the number of segregating sites (\( \theta \)) or average pairwise nucleotide diversity (\( \pi \)) are expected at mutation-drift equilibrium under a standard neutral model. Tajima’s \( D \) statistic quantifies departures from this neutral expectation, and values different from zero suggest the action of nonneutral selective or demographic processes (56). If the gene under study is in mutation-drift equilibrium (i.e., evolving neutrally), the two estimators of DNA polymorphism \( \theta \) and \( \pi \) should cancel out (\( D = 0 \)). Under negative (purifying) selection, new mutations will be deleterious and therefore will not rise in frequency and rare variants will be abundant. Nucleotide diversity is mainly determined by high-frequency mutations and will not be seriously affected by rare mutations. The number of segregating sites is therefore relatively high, which translates into a negative \( D \) statistic. In contrast, under balancing selection, more than one allele is selectively favored and nucleotide diversity will be relatively high, translating into a positive \( D \) statistic.

Tajima’s \( D \) was calculated by using both the total number of segregating sites (\( S \)) and the total number of mutations (\( M \)) (Table 3). The number of mutations, nucleotide diversity (\( \pi \)), and the number of segregating sites corrected for sample size (\( \theta \)) can be found in Table 1. Under the infinite-sites model, the number of segregating (polymorphic) sites equals the total number of mutations (polymorphisms). In several of the gene fragments, however, three or four different polymorphisms were represented at certain polymorphic sites. Especially in the highly diverse \( \text{pilA} \) fragment, this led to a considerable difference between the number of segregating sites and the number of mutations (135 and 184, respectively). Using the number of segregating sites lowers the proportion of rare variants and hence positively influences \( D \). In addition, the \( D \) statistic was calculated for the translated sequence with the program MEGA. Using the relative frequency of particular amino acids in a protein may be a more relevant approach because synonymous substitutions at the DNA level are factored out. We were unable to find previous applications of this version of Tajima’s test in the literature. Tajima’s test requires a random sample from the population, and therefore we used all sequences available (78 for \( \text{csgA} \), \( \text{fibA} \), and \( \text{pilA} \) and 20 for the other genes).

\( D \) values were negative for \( \text{csgA} \), suggesting purifying selection in this gene (although these values were not significant). Smaller negative values were found for \( \text{fibA} \), indicating less constraint by purifying selection. Tajima’s \( D \) values in the other genes varied from \(-0.687 \) to \( 0.983 \) and were not significant (data not shown). In accordance with the balancing selection on \( \text{pilA} \) suggested by the Ewens-Watterson test, positive \( D \) values were found for \( \text{pilA} \) (Table 3). However, values were not significantly different when \( M \) was used. Because selection is likely to operate only on some specific domains of a protein sequence, separate analyses were performed on three regions in this gene fragment. As expected, \( D \) values were lowest in the relatively conserved \( \alpha \)-helix region but were nonetheless positive and therefore did not indicate strong selective constraint. The highest \( D \) values were found in the putative \( D \) region (excluding the two cysteine residues that demarcate this region), providing circumstantial evidence that this loop is indeed protruding from the pilus and involved in interactions with the outside world. The putative \( D \) region is lacking in divergent group VI, and these sequences are thus not included in the calculation of the \( D \) statistic for this region. Large positive values of Tajima’s \( D \) can also arise from a recent reduction in population size. A reduction in population size can eliminate much of the variation present in a population, and not enough time might have passed for new mutations to accumulate. However, this scenario is highly unlikely in the case of \( \text{pilA} \) because of its high degree of polymorphism.

Under neutral evolution, the number of nonsynonymous substitutions per nonsynonymous site (\( dN \)) and the number of synonymous substitutions per synonymous site (\( dS \)) are expected to be equal. Since nonsynonymous substitutions change the primary amino acid structure and therefore might alter protein function, elevated \( dN/dS \) ratios can be interpreted as indicative of positive selection or, alternatively, relaxed selective constraint (34). However, the \( dN/dS \) ratio is not a good measurement of selection when considering very closely re-

### Table 2. Ewens-Watterson test of neutrality

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( csgA )</th>
<th>( fibA )</th>
<th>( pilA )</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of samples</td>
<td>78</td>
<td>78</td>
<td>78</td>
</tr>
<tr>
<td>No. of alleles</td>
<td>13</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Observed ( F ) value</td>
<td>0.244</td>
<td>0.322</td>
<td>0.170</td>
</tr>
<tr>
<td>Expected ( F ) value</td>
<td>0.192</td>
<td>0.414</td>
<td>0.284</td>
</tr>
<tr>
<td>Watterson’s ( F ) value</td>
<td>0.842</td>
<td>0.312</td>
<td>0.043</td>
</tr>
<tr>
<td>Slatkin’s exact ( F ) value</td>
<td>0.397</td>
<td>0.329</td>
<td>0.072</td>
</tr>
</tbody>
</table>

* Two-tailed test, 10,000 replications.
* Significantly different.

### Table 3. Tajima’s test of neutrality

<table>
<thead>
<tr>
<th>Gene</th>
<th>No. of clones</th>
<th>( D ) for nucleotides</th>
<th>( D ) for amino acids</th>
</tr>
</thead>
<tbody>
<tr>
<td>( csgA )</td>
<td>78</td>
<td>(-1.357 )</td>
<td>(-1.357 )</td>
</tr>
<tr>
<td>( fibA )</td>
<td>78</td>
<td>(-0.847 )</td>
<td>(-0.959 )</td>
</tr>
<tr>
<td>( pilA ) whole fragment</td>
<td>78</td>
<td>2.420</td>
<td>0.874</td>
</tr>
<tr>
<td>( pilA ) conserved region</td>
<td>78</td>
<td>1.793</td>
<td>1.400</td>
</tr>
<tr>
<td>( pilA ) variable region</td>
<td>78</td>
<td>2.563</td>
<td>0.712</td>
</tr>
</tbody>
</table>
| \( pilA \) putative \( D \) region | 74 | 3.663 | 1.754 | 4.417 *

* \( S \), segregating sites.
* \( M \), mutations.
* Nucleotide position, bp 1 to 180.
* Nucleotide position, bp 181 to 324.
* Nucleotide position, bp 229 to 276.
* Excluding the four group VI strains that lack the cysteine residues at positions 98 and 115.
* \( P < 0.05 \).
* \( P < 0.01 \).
* \( P < 0.001 \).
lated bacterial sequences because not enough time has passed to remove slightly deleterious nonsynonymous mutations (42). This may explain why, contrary to expectation, the \( dN/dS \) ratio of \( \text{fibA} \) was much higher than that of \( \text{pilA} \) (0.6 versus 0.28). The \( dN/dS \) ratios calculated separately for the different \( \text{pilA} \) regions did correspond in rank to the Tajima \( dN \)/\( dS \) ratios calculated separately for the different \( \text{fibA} \) regions, with the highest value (0.54) for the putative D region (data not shown).

**DISCUSSION**

This study sought to describe the genetic structure of the *M. xanthus* population present in a small soil plot. A total of 78 clones were isolated from 100 soil samples, and 21 genotypes were distinguished among them from a concatemer of the \( \text{cs} \), \( \text{fib} \), and \( \text{pil} \) gene fragments. The genetic variation analyzed in this study was shown to be largely representative of the actual variation present in the sampling plot. Genotypes did not exhibit clustering in the grid, implying that individual *Myxococcus* clones occur in continuous swarms only below the centimeter scale. Comparison of the intraspecific diversity present in this population of *M. xanthus* with diversity estimates for other species must be made cautiously because of differences in species definitions and sampling schemes. However, the intraspecific diversity recovered here does not appear to be higher than that reported in several recent studies on soil bacteria (3, 37, 59).

A recent evolutionary framework on bacterial diversification developed by Cohan (7) describes bacterial populations as complexes of neutrally diversifying clones. Depending on chance and selective pressure, occasionally a mutant better adapted to its environment arises and sweeps through the population by natural selection. The purging of the population through the fixation of this single clone means that the phylogenetic tree is pruned to a single branch. This evolutionary line is called an ecotype. Ecotypes are defined as having different ecological specializations such that selective sweeps within ecotype complexes do not affect other ecotypes.

This model of bacterial evolution finds support in MLST studies on pathogens where the genetic population structure can be best described by a collection of different clonal complexes, i.e., common founder genotypes radiating into an array of neutral offshoot clones (14). However, recent reports on intraspecific variation in free-living marine bacteria describe the occurrence of many closely related genotypes, each occurring at extremely low frequencies (57). These data are not consistent with Cohan's model but rather indicate the buildup of many neutral mutations that are not regularly purged by selective sweeps. Chance, rather than natural selection, is hypothesized to shape the population structure of planktonic bacteria because the combination of extremely low population densities and a patchy distribution of nutrients results in stochastic nutrient encounters (57). In addition, predation might quickly erase any localized dominance of genotypes (57).

The *M. xanthus* population studied here is composed of genotypes that have little nucleotide divergence between them but share very few alleles. Except for two cases, the eBURST algorithm therefore failed to group different clones together. The population structure thus does not resemble the epidemic population structures of pathogens. We cannot exclude the possibility that offshoot clones are missing from our data set because their frequency is below the detection limit of our study. If this were the case, each genotype found here would represent a whole clonal complex from which no other neutral variants were sampled. However, it seems doubtful that offshoot clones are missing from our data set because of their recent removal by selective sweeps. This would require that complete selective sweeps of new superior mutants occur so frequently that neutral offshoot clones are rarely detected.

The structure of this *M. xanthus* population does not closely resemble the picture emerging from studies on the population structure of marine bacteria either. Not all genotypes are rare, and balancing selection appears to be at least partially responsible for the genetic variation observed. Between the extremes of the population structures of pathogens and planktonic bacteria, it is difficult to speculate how selective sweeps influence the genetic diversity of *Myxococcus* clones living in soil. It is plausible that microecological parameters vary sufficiently at the scale of the sampling plot to allow different genotypes to coexist in different niches. A wide range of processes are important in experimental adaptive radiations of microbes, including resource competition (40), interference competition (25), and coevolution with phage (5), and many such forces are likely to shape *M. xanthus* diversity. Perhaps the extremely spatially structured soil habitat even offers clones with deleterious mutations some protection from selective sweeps. Ultimately, it is important to know whether genetic variation is neutral, adaptive, or even deleterious. Toward this end, experiments are under way to characterize phenotypic and behavioral variations in a subset of the clones studied here.

Although the *cs*, *fib*, and *pil* phylogeny served to distinguish clones, it is probably not an accurate reflection of the genomewide evolutionary relationships among the strains especially because the highly diversified *pil* gene presents a strong bias. The random subset of 20 clones for which eight gene sequences were used to construct a concatemer phylogeny provides a more reliable picture of evolutionary relationships. It is apparent that the 10 STs these clones represent are much more closely related to each other than that they are to standard laboratory strain DK1622. The molecular variation summarized by \( \pi \) and \( \theta \) for this local population (Table 1) is at least an order of magnitude lower than that in a recent study on the global population structure of *Pseudomonas syringae* (46). However, strains A66, A75, A88, and A99 (group VI in the *cs*, *fib*, *pil* phylogeny) and strain A98 were not part of this random subset. Based on the *cs*, *fib*, and *pil* phylogenies, group VI and the random subset seem to be roughly as distant from each other as they each are from DK1622. Thus, there seems to be a major phylogenetic cluster (Fig. 6) distinct from a small minority of genotypes (5/78, \( \sim 6\% \)) that is only distantly related to the main cluster.

The presence of a majority of closely related strains is suggestive of a model in which strains within the primary cluster share a largely endemic evolutionary history whereas the distant genotypes represent immigrant genotypes. Myxobacteria are able to form resilient spores that might be carried large distances by migration vectors, which may have been the origin of the group VI clones in this population. The low proportion of distantly related clones could mean that long-range dispersal is not frequent enough to erase evidence of a predom-
inant local population, that most immigrants are maladapted to local conditions and rapidly decrease in frequency upon arrival, or both. Alternatively, the five genotypes outside the primary cluster might have evolved locally but diverged more rapidly than strains in the primary cluster.

The biogeography of free-living bacteria is only beginning to be resolved (32), and case studies in which local diversity seems representative of global diversity (e.g., see references 18 and 55) and case studies indicative of endemic distributions (e.g., see references 6, 35, 38, and 61) have both been reported in the literature. The ectotype model of diversification does not require spatial isolation of clones to permit divergence between them, although Cohan does note that spatial isolation might shelter nascent ecotypes from selective sweeps in the ancestor ectotype until it has accumulated enough ecological adaptations to be fully independent (7). MLST studies of M. xanthus isolates from larger spatial scales are currently under way to further investigate the global biogeography of this species.

The csgA, fibA, and pilA genes showed markedly different patterns of natural selection as inferred from the comparison of allele frequency distributions and the frequency of nucleotide and amino acid sequence polymorphisms relative to the neutral expectation. Evidence for balancing selection was found for the pilin-encoding pilA gene, which has also been detected in the pilin subunit of Neisseria meningitidis (1). Type IV pili in M. xanthus are involved in cell-cell contact and are necessary for social motility and fruiting body formation. Further studies are required to demonstrate functional differences associated with the different pilA alleles. Interestingly, the functional domains in the csgA gene were found to be conserved, suggesting that this signaling gene is unlikely to be responsible for developmental incompatibilities observed in other M. xanthus isolates (17). The fibA gene fragment did not appear to be under either negative or balancing selection.

Strong evidence of past recombination events comes from three clear instances of incongruence displayed by A75, A98, and DK1622 in the csgA, fibA, and pilA phylogenies. We therefore infer that horizontal gene transfer can occur in the species M. xanthus. However, the support for a linkage disequilibrium scenario by the Iq test indicates that recombination appears to be rarer than in many other species (e.g., N. meningitidis) (14). Various additional tests for recombination were employed, but no clear picture emerged of how important recombination events are relative to the accumulation of point mutations in M. xanthus genome evolution. Split decomposition graphs of the individual gene trees were often incongruent with each other, but the split graph of the concatenome sequence was not indicative of recombination. Two phylogenetic methods did not find evidence for recombination, but the maximum chi-squared test did in several cases. Since the latter test is not very likely to produce false positives and is suitable for data sets with low divergence (39), these results should not be dismissed. Sawyer’s runs test provided evidence for recombination in the first half of the pilA gene. Our mixed results highlight the importance of employing multiple tests for recombination to avoid making inferences that may be idiosyncratic to a particular method.

M. xanthus exhibits sophisticated social behaviors throughout its life history, using social motility to communally feed and cooperatively building fruiting bodies upon starvation. This study shows that M. xanthus is surrounded in the soil by a wide range of genetically distinct conspecifics. It can be assumed that sympatric genotypes occasionally come into contact because of either swarming motility or environmental perturbation. It will be of interest to determine the degree to which these genotypes engage in cooperative behavior with one another during swarming and development. A recent study with nine M. xanthus clones isolated from distant global locations showed that clone pairs forced to undergo development in a mixture generally exhibit intense antagonism toward each other (17). In a majority of clone pairings, bidirectional antagonism was observed, with both clones producing significantly fewer spores in mixture than they do in clonal cultures. Since most of the strains isolated in this study are more closely related to each other than those in the global competition study (data not shown), and perhaps have evolved over extended periods within the same patch of soil, social interactions among these clones could be markedly different. The degree of social compatibility among a subset of the Tübingen clones examined here is currently being investigated.

ACKNOWLEDGMENTS

We thank Ed Feil and three anonymous reviewers for helpful comments on the manuscript; Iris Dinkela¨cker, Gerardo Lippert-Viera, and Heike Keller for technical assistance; and Daniel Huson, Stephan Henz, Sara Sarkar, Lori Handley, Joshua Herbeck, Elze Hesse, and Jennifer Hughes for helpful suggestions. M.V. thanks Deborah Charlesworth, John Brookfield, and the BBSRC for organizing the 2004 Summerschool on Molecular Evolution and Diversity and John Brookfield for helpful suggestions.

REFERENCES
