Whole-genome comparison of disease and carriage strains provides insights into virulence evolution in *Neisseria meningitidis*

Christoph Schoen^{*†}, Jochen Blom[‡], Heike Claus^{*}, Anja Schramm-Glück^{*}, Petra Brandt^{§¶}, Tobias Müller^{||}, Alexander Goesmann[‡], Biju Joseph^{*}, Sebastian Konietzny[‡], Oliver Kurzai^{*}, Corinna Schmitt^{*}, Torben Friedrich^{||}, Burkhard Linke[‡], Ulrich Vogel^{*}, and Matthias Frosch^{*}

^{*}Institute for Hygiene and Microbiology, University of Würzburg, Josef-Schneider-Strasse 2, 97080 Würzburg, Germany; [‡]Center for Biotechnology, University of Bielefeld, Universitätsstrasse 27, 33615 Bielefeld, Germany; [§]MWVG Biotech, Anzinger Strasse 7a, 85560 Ebersberg, Germany; and ^{||}Department of Bioinformatics, University of Würzburg, Biocenter, Am Hubland, 97074 Würzburg, Germany

Communicated by Rino Rappuoli, Novartis Vaccines, Siena, Italy, January 10, 2008 (received for review December 5, 2007)

Neisseria meningitidis is a leading cause of infectious childhood mortality worldwide. Most research efforts have hitherto focused on disease isolates belonging to only a few hypervirulent clonal lineages. However, up to 10% of the healthy human population is temporarily colonized by genetically diverse strains mostly with little or no pathogenic potential. Currently, little is known about the biology of carriage strains and their evolutionary relationship with disease isolates. The expression of a polysaccharide capsule is the only trait that has been convincingly linked to the pathogenic potential of *N. meningitidis*. To gain insight into the evolution of virulence traits in this species, whole-genome sequences of three meningococcal carriage isolates were obtained. Gene content comparisons with the available genome sequences from three disease isolates indicate that there is no core pathogenome in *N. meningitidis*. A comparison of the chromosome structure suggests that a filamentous prophage has mediated large chromosomal rearrangements and the translocation of some candidate virulence genes. Interspecific comparison of the available *Neisseria* genome sequences and dot blot hybridizations further indicate that the insertion sequence IS1655 is restricted only to *N. meningitidis*; its low sequence diversity is an indicator of an evolutionarily recent population bottleneck. A genome-based phylogenetic reconstruction provides evidence that *N. meningitidis* has emerged as an unencapsulated human commensal from a common ancestor with *Neisseria gonorrhoeae* and *Neisseria lactamica* and consecutively acquired the genes responsible for capsule synthesis via horizontal gene transfer.

comparative genomics | genome evolution | bacterial capsule | neisserial prophage | IS1655

The meningococcus *Neisseria meningitidis* is a bacterial commensal of the human nasopharynx and for reasons that are still mostly unknown occasionally causes septicemia and meningitis (1). Because of the high lethality and frequent long-term sequelae in survivors, most of the experimental work on pathogenesis has focused on disease strains so far. However, multilocus sequence typing (MLST) (2) has revealed that most cases of disease are caused by only very few clonal complexes of related sequence types (STs) referred to as hypervirulent lineages (2) and that isolates belonging to these hypervirulent lineages are underrepresented in healthy carriers (3–5). In turn, >10% of the healthy population are colonized with *N. meningitidis* and most carriage isolates were found to belong to a bewildering variety of >2,000 different sequence types that rarely, if ever, cause disease.

The genetic basis for the observed virulence differences is still a matter of investigation (6). Because of the lack of an appropriate animal model, most insight in pathogenicity factors of *N. meningitidis* resulted from *in vitro* studies with cell cultures, and thus the *in vivo* relevance for most of these candidate virulence

factors still remains to be validated (7). One of the well established virulence factors in *N. meningitidis* is the polysaccharide capsule (8), which mediates resistance against complement-mediated lysis and opsonophagocytosis (9). Based on the chemical composition and the immunological characteristics of their capsular polysaccharide, meningococci are divided into 13 serogroups with serogroups A–C, W-135, and Y being most frequently associated with human disease (10). However, although carriage isolates are frequently constitutively unencapsulated because of a lack of the genetic island encoding capsule synthesis (11), several carriage isolates express polysaccharide capsules otherwise associated with disease (3–5). These observations suggest that the capsule is necessary, but not sufficient to confer virulence.

For *N. meningitidis*, the genomes of three disease isolates belonging to different hypervirulent lineages have been annotated and published (12–14). Based on these three genomes, comparative genomic hybridization (CGH) studies highlighted the presence of a filamentous prophage termed Nf1 (15) in the genomes of several, but not all hypervirulent lineages (16, 17). Again, its contribution to virulence in these isolates is currently unresolved because it does not code for any known virulence factor.

Here, we provide the genome sequences of three meningococcal carriage strains including the complete genome of an constitutively unencapsulated meningococcal strain belonging to a clonal complex that is frequently recovered from healthy carriers (3). Whole-genome comparisons of the disease and carriage genomes were performed to analyze possible mechanisms of chromosome structure variation, to search for genes that are specific for disease isolates and to gain insight into the evolutionary relationships between the sequenced disease and carriage strains.

Author contributions: C. Schoen, U.V., and M.F. designed research; C. Schoen, J.B., H.C., A.S.-G., P.B., T.M., A.G., and B.J. performed research; C. Schoen, J.B., H.C., A.S.-G., T.M., B.J., S.K., O.K., C. Schmitt, T.F., B.L., and U.V. analyzed data; and C. Schoen, U.V., and M.F. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the DNA Data Base in Japan/European Molecular Biology Laboratory/GenBank databases [accession nos. AM889136 (α14), AM889137 (α153), and AM889138 (α275)].

[†]To whom correspondence should be addressed at: Institute for Hygiene and Microbiology, Josef-Schneider-Strasse 2, Building E1, 97080 Würzburg, Germany. E-mail: cschoen@hygiene.uni-wuerzburg.de.

[¶]Present address: Eurofins Medigenomix GmbH, Fraunhoferstrasse 22, D-82152 Martinsried, Germany.

This article contains supporting information online at www.pnas.org/cgi/content/full/0800151105/DC1.

© 2008 by The National Academy of Sciences of the USA

Table 1. Comparative overview of the sequenced meningococcal genomes

Feature	Strain					
	Z2491*	MC58*	FAM18*	α 14	α 153	α 275
Molecular typing						
Serogroup	A	B	C	cnl	29E	W-135
Sequence type	4	74	11	53	60	22
Clonal complex	ST-4	ST-32	ST-11	ST-53	ST-60	ST-22
Frequency in carriers, % [†]	0.0	5.0	1.0	7.2	3.9	4.6
Frequency in cases, % [‡]	0.0	23.8	18.1	0.0	2.7	0.4
General information						
No. of contigs (>2 kb)	1	1	1	1	87	133
Genome size, bp	2,184,406	2,272,351	2,194,961	2,145,295	2,134,469	2,266,686
G+C content, %	51.8	51.5	51.6	52.0	51.6	51.2
GenBank accession no.	AL157959	AE002098	AM421808.1	AM889136	AM889137	AM889138
Selected repeats						
dRS3	672	689	656	646	710	692
(attccnnnnnnngggaat)						
dRS3/Nf1	291	300	263	252	303	291
(attccgcctgcgcgggaat)						
Coding sequences						
Putative number	1,993	2,063	1,975	1,987	≥1,814	≥1,947
Average CDS length, bp	902	871	918	884	899	870
COGs, %	68	68	ND	78	86	85
Putative pseudogenes	84	92	58	70	ND	ND
Coding area, %	78.9	79.1	80.2	81.8	≥76.4	≥74.7

ND, not determined.

*The numbers given for *N. meningitidis* strains Z2491, MC58, and FAM18 are based on the GenBank accession nos. AL157959, AE002098, and AM421808.3, respectively.

[†]Frequency of clonal complex in 822 carrier isolates as given in ref. 3.

[‡]Frequency of clonal complexes in 525 disease isolates during the time period 2000–2002 in Germany, characterized by the European Meningococcal MLST Centre (Oxford, U.K.).

Results and Discussion

General Features of the Meningococcal Genomes. All three meningococcal strains sequenced in this study belong to different multilocus sequence types and were isolated from healthy carriers during a carriage study performed in 1999/2000 in Bavaria, Germany (3). Strains α 153 and α 275 belong to the serogroups 29E and W-135, respectively, whereas strain α 14 contains the so-called capsule null locus (*cnl*) and thus is incapable of capsule production (11) (Table 1).

The three newly sequenced genomes and the genomes of *N. meningitidis* serogroup A strain Z2491 (13), serogroup B strain MC58 (14), and serogroup C strain FAM18 (12) already available in the databases were used for subsequent analysis (Table 1). In the following, the six strains will be referred to as α 14, α 153, α 275, Z2491, MC58, and FAM18, respectively. All genomes revealed similar benchmark data such as genome size, GC content, or number of predicted genes (Table 1). The extrapolated number of genes belonging to the meningococcal core genome is at least 1,337 [$CI_{0.95} = (1312.8; 1361.2)$] and the number of new genes contributed by each new *N. meningitidis* genome is predicted to be at least 43 [$CI_{0.95} = (38.3; 48.7)$]. This indicates that the pangenome of *N. meningitidis* is open (18) and that new genes continue to be added to the gene pool of *N. meningitidis* any time a new strain is sequenced. Therefore, the construction of a “pan-*Neisseria*” microarray (19, 20) capturing the entire meningococcal gene pool might be an elusive goal.

The Meningococcal Pathogenome. Whole-genome comparisons further showed that there is no consistent difference in the distribution of most of the hitherto-studied candidate virulence genes between carriage and disease isolates. In fact, of the 134 genes listed in a recent compilation of candidate virulence genes found in *N. meningitidis* (19, 21), 115 can also be found in the

carriage isolates [supporting information (SI) Table 2]. Of the remaining 19 genes found exclusively in the disease isolates, only NMA2123 coding for an FrpC-related RTX-family exoprotein in Z2491 was also found to be present in all sequenced disease isolates. However, a number of paralogous proteins to NMA2123 are encoded also in the genomes of the three carriage isolates. Surprisingly, when compared with all of the genes contained in the meningococcal pangenome, a higher proportion of these candidate virulence genes belong to the meningococcal core genome and only very few are restricted to only one strain (Fig. 1).

In a complementary approach, the number of genes that are shared by all disease isolates but are absent from all carriage isolates was calculated for different combinations of disease and carriage isolates. The size of this so-defined core pathogenome declines sharply with increasing number of genomes compared and comprises ≈ 11 genes when all six strains are included (Fig. 1). According to the annotation of Z2491, the genes NMA1792–NMA1799 belong to the filamentous prophage Nf1 (15), which was recently shown to be specific only for certain hypervirulent lineages but absent in others (16, 17). The remaining three genes are all located on the *pha* locus in Z2491. NMA0692 putatively encodes an alternative C-terminal cassette for the effector protein (HrpA) of the meningococcal two-partner secretion system (TPS). However, paralogous genes encoding a complete TPS are present in all carriage isolates analyzed and are likely to be present in all meningococcal clonal lineages (22). Finally, NMA0693 has some sequence similarity to a predicted soluble lytic transglycosylase from *Vibrio vulnificus* (GenBank NP_762383), and NMA0694 encodes a hypothetical protein. Although NMA0692 and NMA0694 were both shown to be expressed in *N. meningitidis* on contact with human epithelial cells *in vitro* (23), it remains questionable whether these genomic differences might also translate into virulence differences.

between different dRS3/Nf1 elements resulting in permanent genomic changes such as the observed translocations or deletions of dRS3/Nf1-enclosed genes. Because many of these genes code for surface proteins involved in the interaction of bacteria with host cells (12), their chromosomal translocation could in turn lead to the observed virulence differences between hypervirulent lineages and carriage isolates because of an altered expression of the affected genes. Future studies should therefore assess whether gene or protein expression profiles are modulated by such chromosomal translocations.

In face of the high selective pressure for inactivation of transposases in the meningococcal genomes (29), the low number of inactivating mutations (SI Fig. 6) together with the inferred high mobility of IS1655 provide evidence for a quite recent infection of the meningococcal genomes with this element. This conclusion is also supported by the remarkably low intergenomic and intragenomic sequence diversity of IS1655 (SI Table 4) that was shown not to be the result of gene conversion events (30). Taken together, these data indicate an evolutionarily quite recent population bottleneck because of, e.g., a change in the lifestyle of *N. meningitidis* or its ancestor such as specialization to novel human-related niches and/or mode of transmission.

mon split with *N. gonorrhoeae* and *N. lactamica*, indicating that among the meningococcal strains compared $\alpha 14$ is closest to *N. gonorrhoeae* and *N. lactamica* despite the inferred high rate of horizontal gene transfer (HGT). Also, with respect to its IS profile, $\alpha 14$ is the meningococcal strain most similar to *N. lactamica* and *N. gonorrhoeae*, respectively, among the genomes compared (Fig. 3).

Schoen et al.

previous observations of HGT from *Haemophilus influenzae* being also a member of the *Pasteurellaceae* and inhabitant of the human airways to *N. meningitidis* (35). Therefore, the encapsulated and thus potentially pathogenic strains of *N. meningitidis* might have evolved from an unencapsulated ancestor by horizontal acquisition of the *cps* locus from other bacteria residing in the human nasopharynx (36).

Conclusion and Outlook. Our results from whole-genome comparisons suggest an origin of the encapsulated meningococcal strains from an unencapsulated common ancestor with *N. gonorrhoeae* and *N. lactamica*. Historically, meningococcal disease is believed to have emerged quite recently with definitive descriptions dating only from 1805 in Europe and North America and 1905 in Africa (37). The genes required for capsule synthesis were acquired from other bacterial species residing in the human nasopharynx via HGT only very recently, and our data suggest that the sources were members of the family *Pasteurellaceae*. This proposed recent emergence of *N. meningitidis* in turn explains the highly similar genomic makeups of disease and carriage isolates. With the phage Nf1 possibly modulating the virulence of the infected strains, differences in the pathogenic potential between carriage and disease isolates might thus be influenced by small genetic differences in genes from the core genome. One prime candidate would be variation in the repeat tracts of phase variable genes associated with host cell interactions (12), which will be detectable as single nucleotide polymorphisms (SNPs) in larger collections of genome sequences from freshly isolated carriage and disease strains. Whole-genome association mapping of resulting SNP haplotypes with the propensity of a strain to cause disease might reveal the anticipated genetic differences. Novel high-throughput sequencing technologies (38) will enable the sequencing of the large number of further strains and species required (39) to test the proposed evolutionary scenario and the hypotheses on the emergence of virulence in *N. meningitidis*.

Materials and Methods

Genome Sequencing and Assembly. The three sequenced *N. meningitidis* strains α 14, α 153, and α 275 were chosen from a collection of 830 meningococcal isolates obtained during a carriage study in 8,000 children and young adults in Bavaria, Germany, in 1999/2000 (3). Genome sequences were generated by whole-genome shotgun sequencing. The single contig of strain α 14 was assembled from 32,059 reads (giving a $9.38\times$ coverage) from pGEM-Teasy (Promega) clones with 2.0- to 3.0-kb inserts. The draft genomes of strains α 153 and α 275 were sequenced to $8\times$ coverage from 29,283 reads for strain α 275 and from 26,630 reads for strain α 153, respectively, and the resulting non-overlapping contigs were pasted together into a pseudochromosome in random order with the sequence 5'-CTAGCTAGCTAG-3' used as spacer that generates a stop codon in all six reading frames.

Genome Annotation. Automated annotations of the assembled genome sequences were carried out with the genome annotation system GeneDB2 (40), followed by manual curation for strain α 14. For the comparative analysis of IS content, the genome sequences were searched with TBLASTN (41) using the neisserial ISs deposited in the ISfinder database (42) as query sequences. Based

on the score ratio of the best hit in the genome versus the self hit, ISs and strains were clustered based on the correlation distance using Cluster 3.0 (43).

Whole-Genome Alignments. The annotated genomes of *N. meningitidis* strains Z2491 (AL157959) (13), MC58 (AE002098) (14), and FAM18 (AM421808.3) (12) were used for whole-genome alignments. Pairwise BLASTN alignments of genome sequences were visualized by using the Artemis Comparison Tool Release 5 (44). Multiple whole-genome alignments were performed by using Mauve 2.0, which has been particularly designed for the identification and alignment of conserved genomic DNA in the presence of rearrangements and HGT (33).

Estimation of Core Genome and Pangenome Sizes. Orthologous proteins were operationally identified as reciprocal best matches in BLASTP and TBLASTN comparisons, respectively. A coding sequence was considered to be orthologous to the query sequence if there was $>50\%$ amino acid sequence identity over at least 50% of the query sequence length. The core and pangenome sizes were estimated by nonlinear regression as described in ref. 18. For analyzing the meningococcal pathogenome, the results of BLASTP comparisons were combined with genomic BLASTN alignments to assess the synteny of putative orthologs and to reduce the number of genes falsely detected as pathogen-specific because of a missing annotation in the carriage strains.

Dot Blot Hybridization. For DNA/DNA dot blot hybridizations, either 20 μ l of suspensions of 10^{10} colony forming units per ml or 200 ng of chromosomal DNA were spotted onto nylon membranes (Macherey-Nagel). Dot blot hybridizations were performed by using digoxigenin-labeled probes as described in ref. 45. The following oligonucleotides (Sigma-Aldrich) were used to generate an IS1655 probe by PCR: HC277R (5'-CGC CAA GAC AAA AGC AAC GG-3') and HC273 (5'-ATA CGC CAC AAT AGC TCA GC-3').

Genome-Based Phylogenetic Analyses. For phylogenetic analyses, the genome sequences of *N. gonorrhoeae* FA1090 (AE004969.1) and *N. lactamica* (www.sanger.ac.uk/Projects/Microbes/) were used as outgroups.

SPRING (46) was used for the calculation of rearrangement and breakpoint distances, respectively, from the order of locally collinear blocks (LCBs) being present in all genomes compared. Based on the resulting distance matrices, SplitsTree4 (47) was subsequently used for a neighbor-net reconstruction (48). The robustness of the resulting network topology was confirmed with respect to variation of the parameter settings and the number of genomes compared. In addition, a jackknife resampling method was applied that consisted of the random omission of 10% of the LCBs from each of the LCB order files compared as generated by SPRING.

For the construction of a neighbor-net based on the differential distribution of COGs in the neisserial genomes, genes with COG assignment were selected based either on their primary annotation or on PSI-BLAST hits in the COG database with an E value cutoff of $1E-10$. All-against-all reciprocal BLASTP comparisons were performed and a corresponding presence-absence-matrix based on the differential distribution of 2,633 COG annotated proteins was generated. For each genome pair, a genome distance based on Dollo parsimony (49) was computed by using SplitsTree4 and a network was generated based on the neighbor-net algorithm. To assess the statistical robustness of the resulting network, a bootstrap analysis was performed with 10,000 resampling steps.

ACKNOWLEDGMENTS. We gratefully acknowledge Julian Parkhill for the permission to use the *N. lactamica* genome sequence for the phylogenetic comparisons, Martin Maiden for helpful critique on phylogenetic inference in *N. meningitidis*, and Gabriele Gerlach for critical reading of the manuscript. The work on the meningococcal genomes was supported by the Bundesministerium für Bildung und Forschung in the context of the PathoGenoMik and PathoGenoMik-Plus funding initiative.

1. Rosenstein NE, Perkins BA, Stephens DS, Popovic T, Hughes JM (2001) Meningococcal disease. *N Engl J Med* 344:1378–1388.
2. Maiden MC, et al. (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 95:3140–3145.
3. Claus H, et al. (2005) Genetic analysis of meningococci carried by children and young adults. *J Infect Dis* 191:1263–1271.
4. Jolley KA, et al. (2000) Carried meningococci in the Czech Republic: A diverse recombining population. *J Clin Microbiol* 38:4492–4498.
5. Yazdankhah SP, et al. (2004) Distribution of serogroups and genotypes among disease-associated and carried isolates of *Neisseria meningitidis* from the Czech Republic, Greece, and Norway. *J Clin Microbiol* 42:5146–5153.
6. Jolley KA, Wilson DJ, Kriz P, McVean G, Maiden MC (2005) The influence of mutation, recombination, population history, and selection on patterns of genetic diversity in *Neisseria meningitidis*. *Mol Biol Evol* 22:562–569.

7. Bourdoulous S, Nassif X (2006) in *Handbook of Meningococcal Disease*, eds Frosch M, Maiden MC (Wiley-VCH, Weinheim, Germany), pp 257–272.
8. Elias J, et al. (2006) Spatiotemporal analysis of invasive meningococcal disease, Germany. *Emerg Infect Dis* 12:1689–1695.
9. Frosch M, Vogel U (2006) in *Handbook of Meningococcal Disease*, eds Frosch M, Maiden MC (Wiley-VCH, Weinheim, Germany), pp 145–162.
10. Maiden MC, Caugant DA (2006) in *Handbook of Meningococcal Disease*, eds Frosch M, Maiden MC (Wiley-VCH, Weinheim, Germany), pp 17–35.
11. Claus H, Maiden MC, Maag R, Frosch M, Vogel U (2002) Many carried meningococci lack the genes required for capsule synthesis and transport. *Microbiology* 148:1813–1819.
12. Bentley SD, et al. (2007) Meningococcal genetic variation mechanisms viewed through comparative analysis of serogroup C strain FAM18. *PLoS Genet* 3:e23.
13. Parkhill J, et al. (2000) Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* 404:502–506.