

RESEARCH ARTICLE

# Complete Genome Sequence and Comparative Genomics of a Novel Myxobacterium *Myxococcus hansupus*

Gaurav Sharma, Tarun Narwani, Srikrishna Subramanian\*

CSIR-Institute of Microbial Technology, Sector-39A, Chandigarh, India

\* [krishna@imtech.res.in](mailto:krishna@imtech.res.in)



**OPEN ACCESS**

**Citation:** Sharma G, Narwani T, Subramanian S (2016) Complete Genome Sequence and Comparative Genomics of a Novel Myxobacterium *Myxococcus hansupus*. PLoS ONE 11(2): e0148593. doi:10.1371/journal.pone.0148593

**Editor:** Feng Gao, Tianjin University, CHINA

**Received:** September 10, 2015

**Accepted:** January 20, 2016

**Published:** February 22, 2016

**Copyright:** © 2016 Sharma et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The complete genome of *Myxococcus* sp. *mixupus* was deposited in GenBank under the accession number CP012109 with Bio Project id PRJNA167109.

**Funding:** This work is supported by a project "Expansion and modernization of Microbial Type Culture Collection and Gene Bank (MTCC) jointly supported by Council of Scientific and Industrial Research (CSIR) Grant No. BSC0402 and Department of Biotechnology (DBT) Govt. of India Grant No. BT/PR7368/INF/22/177/2012". GS acknowledges CSIR for research fellowship. TN is supported by the CSIR network program GENESIS-CSIR-BSC-0121.

## Abstract

Myxobacteria, a group of Gram-negative aerobes, belong to the class  $\delta$ -proteobacteria and order Myxococcales. Unlike anaerobic  $\delta$ -proteobacteria, they exhibit several unusual physiogenomic properties like gliding motility, desiccation-resistant myxospores and large genomes with high coding density. Here we report a 9.5 Mbp complete genome of *Myxococcus hansupus* that encodes 7,753 proteins. Phylogenomic and genome-genome distance based analysis suggest that *Myxococcus hansupus* is a novel member of the genus *Myxococcus*. Comparative genome analysis with other members of the genus *Myxococcus* was performed to explore their genome diversity. The variation in number of unique proteins observed across different species is suggestive of diversity at the genus level while the overrepresentation of several Pfam families indicates the extent and mode of genome expansion as compared to non-Myxococcales  $\delta$ -proteobacteria.

## Introduction

Myxobacteria are Gram-negative  $\delta$ -proteobacteria [1, 2] which are mostly aerobic with some notable exception such as *Anaeromyxobacter* [3]. A peculiar trait of Myxobacteria is their social communication within swarms [2] wherein numerous cell-cell interactions define some of their physiological attributes such as gliding motility [4], fruiting body formation [5], biofilm production, [6] and hunting prey characteristics [7]. Myxobacteria display gliding movement like cyanobacteria and flexibacteria, however, the process is more distinct [8] exhibiting two different types of motilities viz., adventurous and social. Adventurous motility (A) is attributed to a single cell while coordinated movement by a swarm is termed as social motility (S) [9]. Under starvation conditions, Myxobacteria form complex fruiting bodies composed of dormant myxospores, analogous to stalk formation in higher-order fungi [10, 11]. Owing to their complex life cycle, Myxobacteria contain many proteins involved in signal transduction pathways and transcriptional regulation [2]. These proteins help in regulating cell-cell communication and coordinate social motility and fruiting body formation. Besides these unique physiological properties, the relatively large genome size (4.5–15 Mbp) is a characteristic feature of order Myxococcales that distinguish it from other  $\delta$ -proteobacteria (typically 2–7 Mbp)

**Competing Interests:** The authors have declared that no competing interests exist.

[5, 12, 13]. The smallest member of the order Myxococcales is *Vulgatibacter incomptus* DSM 27710 with a genome size of 4.35 Mbp (CP012332.1) followed by *Anaeromyxobacter* with a genome size of ~5 Mbp, which is comparable in size to other non-Myxococcales  $\delta$ -proteobacteria [3]. However, the myxobacterium *Sorangium cellulosum* So0157-2 (14.78 Mbp) [12] is one of the largest genomes among the bacterial clade known till date. The expansion of genome size in Myxococcales is reported to be widespread in all constituent families like; Myxococaceae, Cystobacteraceae, Kofleriaceae and Polyangiaceae [5]. Expansion of a genome indicates increased complexity, influenced by environmental factors and occurrence of genetic events such as duplication and integration of foreign genes via horizontal gene transfer [14]. A large number of duplicated proteins found in Myxobacterial genomes has been suggested to help it adapt to diverse habitats and help in its complex life cycle [15].

Here we report a novel Myxobacterial genome which was found growing as a contaminant in a culture plate of *Chondromyces apiculatus* DSM436 procured from the DSMZ culture collection. We have assembled the complete genome of *Myxococcus hansupus* (named after Dr. Hans Reichenbach and herein referred to as *M. hansupus* or *Mh*) and performed its comparative genome analysis with all available genomes in the genus *Myxococcus* viz., *M. xanthus* DK1622, *M. fulvus* HW-1, *M. stipitatus*, *M. xanthus* DZ2 and *M. xanthus* DZF1 [15–19]. Analysis of these genomes was carried out to help understand the extent of conservation and variability of proteomes in these closely related organisms.

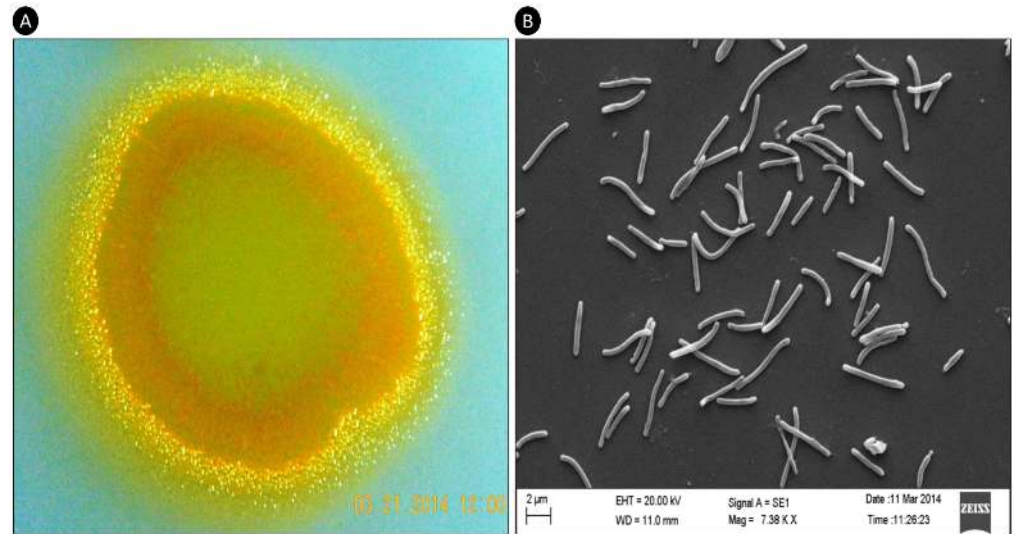
## Material and Methods

### Culturing and DNA isolation of *M. hansupus*

*M. hansupus* was purified from a contaminant on the culture plate of *Chondromyces apiculatus*, procured from Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) culture collection as strain number DSM-436. It was grown on VY/2 agar and SP agar medium plates and is reddish yellow in color (Fig 1A). The swarms were soft and slimy, evenly spread as a film on the agar surface, unlike the *Chondromyces* whose swarms imprint shallow depressions and ridges on the agar [20]. Under scanning electron microscope they look rod-shaped (Fig 1B). Whole genomic DNA was isolated from the pure culture using ZR Fungal/Bacterial DNA MicroPrep™. 16S rRNA sequencing of the isolated DNA was performed using universal bacterial primers at our in-house Sanger sequencing facility. The strain was named *M. hansupus* and is maintained in our laboratory as an actively growing culture.

### Whole genome sequencing and assembly of *M. hansupus*

Sequencing was performed on a Pacific Biosciences RSII instrument at the Genome Quebec Innovation Center, McGill University, Montréal (Québec), Canada. SMRTbell library was constructed with 10  $\mu$ g whole genomic DNA using a 20 kb Template Preparation method and BluePippin™ Size Selection. The library was then loaded onto two single-molecule real-time (SMRT) cells and sequenced using P6 polymerase and C4 chemistry (P6C4) with 180-minute movie times. Sequencing yielded a total of 145,073 reads with a mean read length of 10,730 bp and 1,556,757,303 bp with an estimated coverage of 138 $\times$ . *De novo* assembly was carried out using the hierarchical genome assembly process (HGAP) protocol from SMRT Analysis v2.0, including consensus polishing with Quiver [21]. Gene prediction and functional annotation were performed by Rapid Annotation using Subsystem Technology (RAST) [22]. RNAmmer 1.2 and tRNAscan-SE-1.23 were used to predict rRNA and tRNA genes [23, 24]. The complete genome was used as a reference to determine the putative methylome of *M. hansupus* genome using base modifications and enriched motifs identification protocol of the SMRT portal.



**Fig 1. Colony and cellular morphology.** **A.** Colony morphology of *M. hansupus* swarm with 1x magnification; **B.** Cellular morphology as visualized with Scanning Electron Microscopy (SEM) with 7,380x magnification.

doi:10.1371/journal.pone.0148593.g001

### Data source for comparative genome analysis

Besides the complete genome of *M. hansupus* (*Mh*; CP012109), genome sequences of *M. xanthus* DK1622 (*MxDK1622*; NC\_008095.1) [25]; *M. fulvus* HW-1 (*Mf*; NC\_015711.1) [26]; *M. stipitatus* [*Ms*; NC\_020126.1] [16]; *M. xanthus* DZ2 (*MxDZ2*; AKYI00000000) [18] and *M. xanthus* DZF1 (*MxDZF1*; AOBT00000000) [17] were obtained from NCBI for this study. *MxDK1622*, *Mf*, and *Ms* are complete genomes while *MxDZF1* and *MxDZ2* are draft assemblies. For all these genomes, gene prediction and functional annotation were done using Rapid Annotation using Subsystem Technology (RAST). We also analyzed the replication origin in *M. hansupus* and compared it with those identified in other *Myxococcus* genomes. Complete genomes of *M. hansupus* and other genus *Myxococcus* members were subjected to BLASTn against oriC sequences available at the DoriC database [27, 28].

### Gene identification and reannotation of myxobacterial genomes

In order to have similar annotations for comparative genomics, and to identify annotation inconsistencies, we subjected all the aforementioned genomes to different gene calling and annotation protocols. Various annotation pipelines like RAST [22], GLIMMER [29], xBASE [30] were used in this study using a minimum gene length of 100 bp. Annotated protein-sets from all pipelines were mapped to each other along with the original dataset available in NCBI using BLASTp [E-value cutoff of  $1e^{-5}$ ]. For all genome and pipeline combinations, percentage mapping within each annotation combination was calculated.

### Phylogenetic analysis of *M. hansupus* using 16S rRNA and housekeeping proteins

16S rRNA sequences from the genus *Myxococcus* were extracted from NCBI. Forty *Myxococcus* 16S rRNA sequences along with five out-group sequences (one from each of *Corrallococcus*, *Cystobacter*, *Anaeromyxobacter*, *Sorangium* and *Bdellovibrio* groups) were aligned using the ClustalW module of BIOEDIT sequence alignment tool (version 7.1.3.0) [31]. Post alignment, all the

gaps were excluded and the resulting alignment was used in MEGA 6.06 [32] to generate a maximum likelihood tree [model: Tamura 3-param; bootstrap: 100]. Using the Neighbor-Joining method, initial tree(s) for the heuristic search were obtained and pairwise distance matrix was estimated using the Maximum Composite Likelihood approach. Newick notation of the tree was extracted and used as input in iTOL [33] to generate an interactive phylogenetic tree. Further, genus *Myxococcus* phylogeny was studied using conserved housekeeping genes. Twenty-eight housekeeping genes (*dnaG*, *frr*, *nusA*, *pgk*, *pyrG*, *rplC*, *rplD*, *rplE*, *rplF*, *rplK*, *rplL*, *rplM*, *rplN*, *rplP*, *rplS*, *rplT*, *rpmA*, *rpoB*, *rpsB*, *rpsC*, *rpsE*, *rpsI*, *rpsJ*, *rpsK*, *rpsM*, *rpsS*, *smpB* and *tsf*) [34] were found to be conserved in the complete and draft genomes under investigation (six *Myxococcus* genera, four neighbor genera and one non-*Myxococcales*  $\delta$ -proteobacteria genus, *Bdellovibrio*). Protein sequences of these housekeeping genes were extracted from each genome and concatenated. These concatenated sequences were aligned using ClustalW module of BIOEDIT sequence alignment tool. Gaps were excluded post alignment and the resulting alignment was used as an input in MEGA 6.06 [32] to generate Maximum Likelihood tree [model: Jones-Taylor-Thornton (JTT) matrix; bootstrap: 100]. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distance estimated using a JTT model.

### Orthology, homology and protein clustering study

Orthology was predicted among protein datasets of the six genomes using the Reciprocal Best Hits (RBH) BLAST approach of Proteinortho [35] with an E-value cutoff of  $1e^{-5}$ , minimum query coverage of 50% and minimum identity of 35%. The program first performs an all-against-all BLASTp alignment and then defines putative orthology-pairs based on reciprocal BLAST scores. A cluster is defined by the presence of a protein in at least two genomes. NCBI BLAST+ (v 2.2.26+) was used throughout the study [36].

Homology at protein level was studied among all genomes. Protein dataset from each genome was mapped against the other using BLASTp with an E-value cutoff of  $1e^{-5}$ , minimum query coverage of 50% and minimum identity of 35%. A binary approach was followed to analyze the occurrence of each protein in different genomes. A binary map was generated based on the count of each protein's presence/absence in various genome combinations. For clustering analysis, protein dataset from each genome was mapped against the same using BLASTp with an E-value cutoff of  $1e^{-5}$ , minimum query coverage of 50% and minimum identity of 35%. The filtered dataset for each genome was used to identify the clusters sharing all possible homologs. The mummer program from MUMmer 3.5 suite was used to generate alignment between genome pairs with a minimum alignment length cutoff of 50 bp and mummerplot was used to generate synteny plots [37].

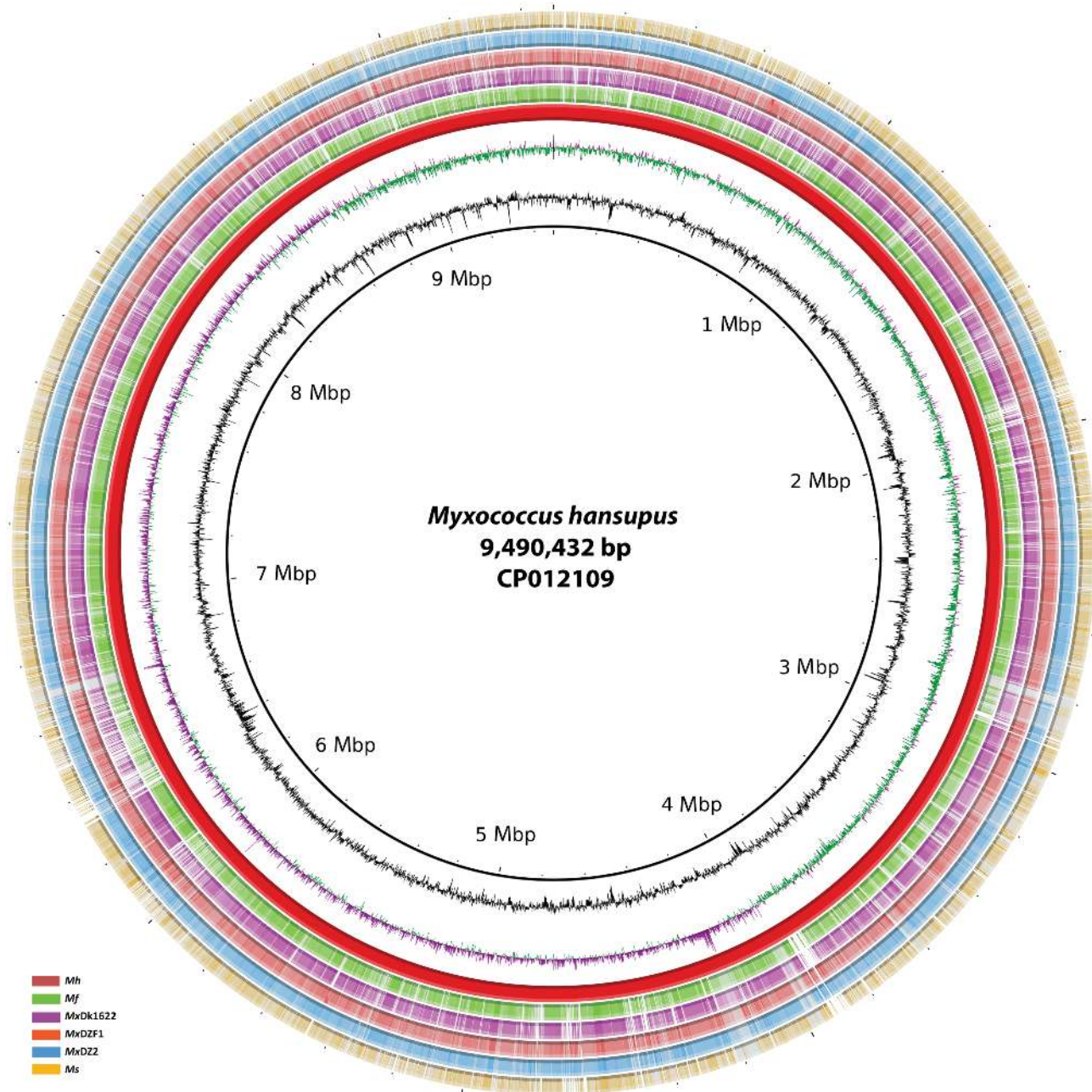
### Pfam domain analysis and core family identification

The proteome of the six *Myxococcus* members and other order Myxococcales members were scanned against the Pfam-A v 28.0 database [38] with an E-value threshold of  $1e^{-5}$  to identify functional domains and other known sequence motifs using hmmscan program of HMMER suite (<http://hmmer.janelia.org/>) [39]. The distribution of Pfam domain families among all genomes was analyzed.

## Results and Discussion

### Genome assembly and annotation

*M. hansupus* genome was assembled as a single chromosome of 9,490,432 nucleotides (Fig 2). The GC content is 69.2% and is comparable to other Myxobacteria [5, 15]. The RNA analysis



**Fig 2. Circular representation of the *M. hansupus* complete genome.** Circles (from inside to outside) **1 and 2** (GC content; black line and GC skew; magenta and green lines), **circle 3** (*M. hansupus*; red circle); **circle 4** (mapped *Myxococcus fulvus* HW-1 genome with *M. hansupus* genome; green circle); **circle 5** (mapped *Myxococcus xanthus* DK1622 genome with *M. hansupus* genome; purple circle); **circle 6** (mapped *Myxococcus xanthus* DZF1 genome with *M. hansupus* genome; Orange circle); **circle 7** (mapped *Myxococcus xanthus* DZ2 genome with *M. hansupus* genome; blue circle); **circle 8** (mapped *Myxococcus stipitatus* genome with *M. hansupus* genome; yellow circle). BRIG 0.95 was used to build the circular representation [53]. Mapping studies were done using BLASTn with an E-value cut-off  $1e^{-5}$ .

doi:10.1371/journal.pone.0148593.g002

of the genome reported four rRNA operons (5S-16S-23S) and 67 amino acyl-tRNA synthetase genes for all twenty amino acids. RAST based annotation helped identify 7,753 coding genes, out of which 4,953 proteins (63.89%) were functionally annotated while the remaining (36.11%) are hypothetical proteins. The coding density of the genome is 88.87% with an

average gene length of 1088 bp. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession CP012109. The genomic features are listed in [Table 1](#).

In *M. hansupus* the replication origin was identified at 8,613,829–8,614,077 bp and the corresponding *dnaA* gene was located downstream of the replication origin at 8,646,592–8,645,240 bp. It shows maximum similarity with *M. fulvus* replication origin (ORI94030396, 365 bp) with an E-value of 9e-68, 89% identity, and 68% coverage and shows 85% sequence identity with 76% length coverage and an E-value of 1e-29 with *M. xanthus* DK1622 (ORI92210206, 247 bp).

Putative methylome of the *M. hansupus* was identified which revealed m6A based methylation in motifs CCAAGGC (82.4% motifs), CTACNNNNNTGG (79.2% motifs), CCANNNNNGTAG (78.1% motifs), SCCCGCA (53.3% motifs), WCCCGCAWG (45.2% motifs) and GATC (31.9% motifs) at 4th, 3rd, 3rd, 7th, 7th and 2nd positions respectively. We identified type I methylases (specific to Adenine) involved in Type I R&M system (AKQ64130, AKQ67990, AKQ68170 and AKQ68203; having N6\_Mtase (PF02384)) but Type II methylases corresponding to Type II R&M systems could not be identified. We also found m4C methylation in motif GCGSYDTY (in only 8.3% motifs) at C2 position. We could not identify corresponding N4-methylcytosine (m4C) methylase while other methylases having Pfam domain N6\_N4\_Mtase (PF01555), which function as both N-4 cytosine-specific and N-6 Adenine-specific DNA methylases, were identified in *M. hansupus* genome (AKQ64825, AKQ65130, AKQ65131, AKQ66512 and AKQ67727). These findings are in accordance with the REBASE database of DNA restriction and modification enzymes [40].

**Table 1. Assembly statistics for *M. hansupus*.**

Organism name	<i>Myxococcus hansupus</i>
Sequencing data	PacBio P6C4 chemistry sequencing
Total Reads	145,073
Number of Bases	1,556,757,303 bp
Mean Read Length	10,730 bp
Average Reference Coverage	138.05 X
Bio-Project Number	PRJNA167109
NCBI Accession number	CP012109
Genome size	9,490,432 bp
GC content	69.2%
Chromosome	1
CDS	7,753
% Coding density	88.87
CDS from (+) strand	3,909
CDS from (-) strand	3,844
Max. CDS length	32,543 bp
Mean CDS length	1,088 bp
Hypothetical proteins	2,800
Hypothetical proteins %	36.11
tRNA	79
Genes with Pfam domains	5,409 (69.77%)
Genes with COG domains	5,650 (72.88%)
Genes with TIGR domains	3,686 (47.54%)

doi:10.1371/journal.pone.0148593.t001

**Table 2. General features of genus *Myxococcus* genomes as annotated by RAST.**

Organism name	<i>M. fulvus</i> HW-1	<i>M. stipitatus</i> DSM 14675	<i>M. xanthus</i> DK 1622	<i>M. hansupus</i>	<i>M. xanthus</i> DZF1	<i>M. xanthus</i> DZ2
Bio Project Number	PRJNA68443	PRJNA186549	PRJNA58003	PRJNA167109	PRJNA199916	PRJNA199464
NCBI Accession	NC_015711.1	NC_020126.1	NC_008095.1	CP012109	AOBT00000000	AKYI00000000
Chromosome/Contigs	1	1	1	1	75	87
GC %	70.6	69.2	68.9	69.2	68.8	68.9
Size (Mbp)	9	10.35	9.14	9.49	9.28	9.27
Genes [RAST]	7,433	8,293	7,524	7,753	7,700	7,689
Coding% [RAST]	89.27	90.22	89.52	88.87	89.12	88.90
Hypothetical proteins (%)	3,002 (40.39)	3,567 (43.01)	2,942 (39.10)	2,800 (36.11)	3,092 (40.16)	3,266 (42.46)
tRNA [RAST]	76	85	75	79	61	61

doi:10.1371/journal.pone.0148593.t002

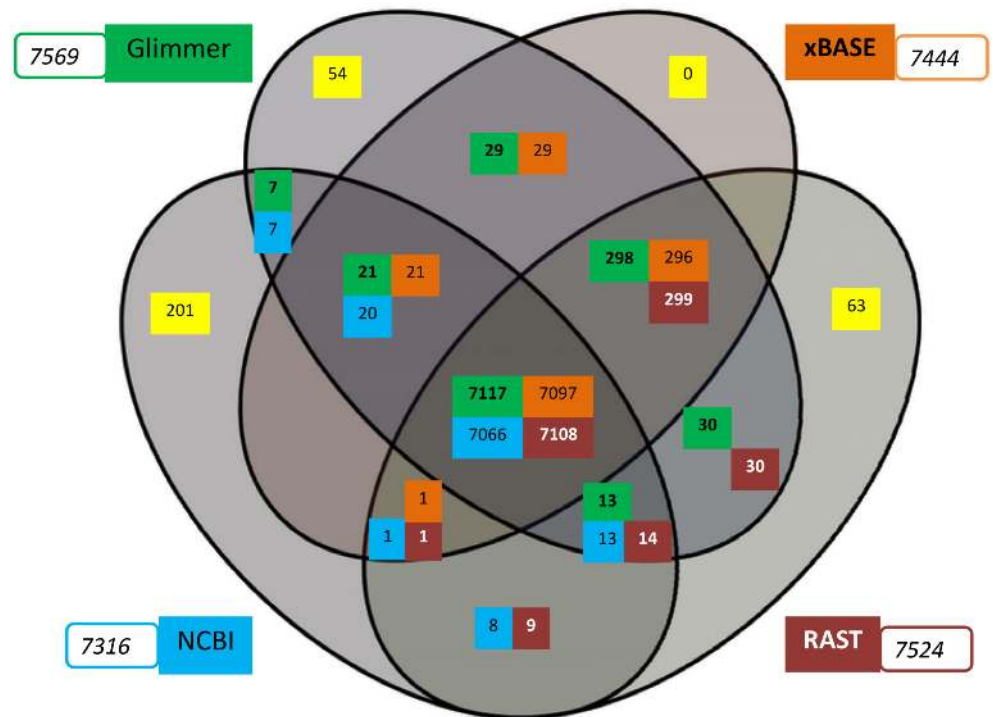
### Genomic overview of the *Myxococcus* clade

At the time of this study, five genomes were available in the genera *Myxococcus* viz., *M. xanthus* DK1622 [25] (*MxDK1622*), *M. fulvus* HW-1 [26] (*Mf*), *M. stipitatus* [16] (*Ms*), *M. xanthus* DZ2 [18] (*MxDZ2*) and *M. xanthus* DZF1 [17] (*MxDZF1*) (Table 2). Among the *M. xanthus* strains, *MxDZ2* is known to be the parent strain of both *MxDK1622* and *MxDZF1* [41]. Including *M. hansupus*, these genomes represent the *Myxococcus* clade belonging to family Cystobacteraceae under suborder Cystobacterineae of the order *Myxococcales*. Among these *MxDZF1* and *MxDZ2* are draft assemblies with 75 and 87 contigs respectively. Noticeably, genome size of non-*Myxococcales* *Deltaproteobacteria* members varies in the range of 2 to 7 Mbp which is relatively smaller as compared to the *Myxococcus* genomes which vary between 9 Mbp to 10.35 Mbp (Table 2). Such genome expansion has been attributed to gene duplication, gene rearrangement, and horizontal gene transfer events [19, 42]. All these strains are reported to undergo developmental program leading to fruiting body formation and can perform gliding motility. Owing to such atypical characteristics, these bacteria pursue a complex life cycle that requires a wide range of proteins functioning coherently. The increased protein content in order *Myxococcales* ranging from 7400–8200 as compared to 4000–5000 in non-*Myxococcales*  $\delta$ -proteobacteria is perhaps in part involved in regulatory functions as reported in earlier studies [15, 25].

### Comparison of annotation pipelines

Various optimized genome annotation pipelines such as RAST [22], Glimmer [29], xBASE [30], PGAAP, JCVI, IGS, and IMG-ER have been used to predict and annotate genes. As the genomes compared here have been annotated using different annotation pipelines at different time points by various groups, therefore we have reannotated the genomes in order to have a consistent and updated annotation of all genomes [43]. Comparative annotation studies were performed to map the annotations with each other to ensure that none of the coding regions in genomes is missed out due to algorithm limitations. We have used the annotation pipelines of RAST, xBASE and Glimmer and also compared these annotations with the original datasets available at NCBI. Comparative mapping studies of all datasets illustrate that annotations using RAST server, GLIMMER, xBASE and original dataset (from NCBI) are comparable to each other with ~97% of proteins being shared amongst them (data not shown).

Taking *MxDK1622* genome as a model, we analyzed the results from different pipelines in order to compare the robustness of the annotation statistics (Fig 3). It was observed that



**Fig 3. Venn diagram of comparative annotations of *Myxococcus xanthus* DK1622 using RAST, Glimmer, xBASE and the original NCBI annotation.** All genome annotations were mapped to each other using BLASTp [E-value cutoff:  $1e^{-5}$ ]. The diagram depicts the homologous proteins shared between two or more annotations (overlapping area) along with unique proteins (yellow shade). RAST, Glimmer, xBASE and the original NCBI annotations are shown in brick red, green, orange and blue colors respectively. The number of annotated proteins using the respective annotation pipeline is shown in the box.

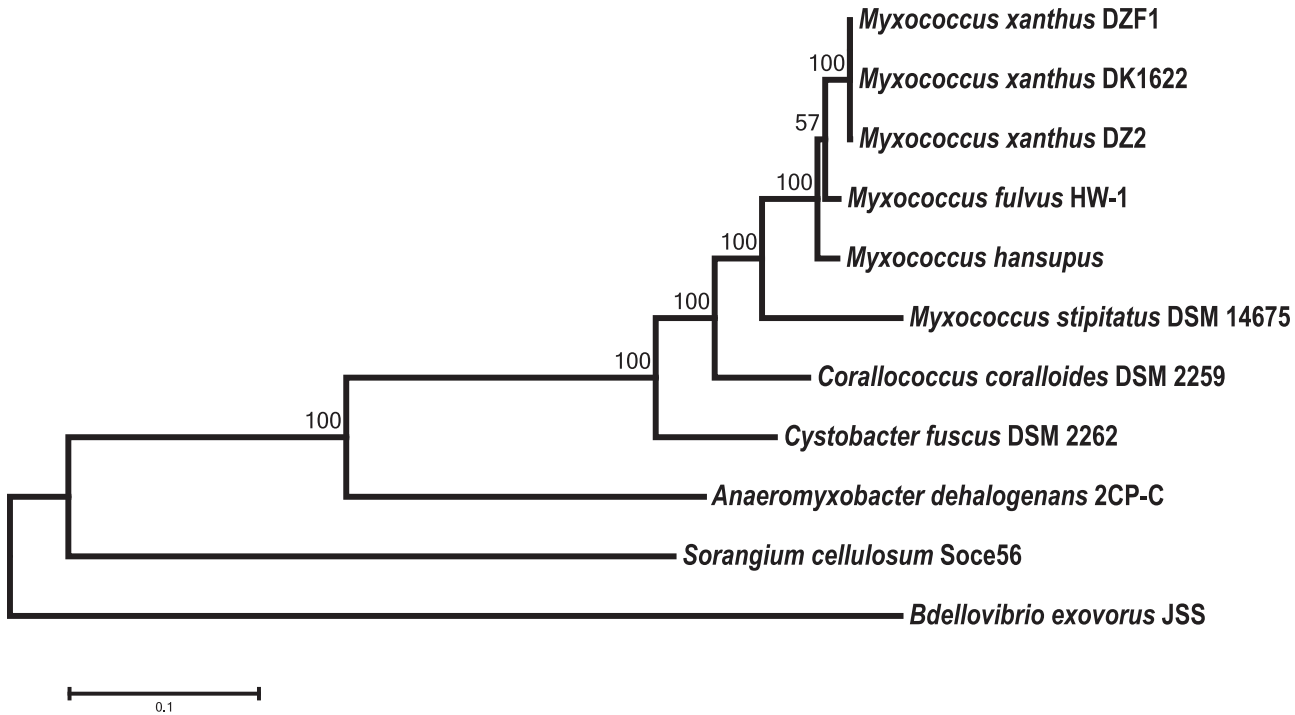
doi:10.1371/journal.pone.0148593.g003

annotations from different pipelines are comparable owing to the similar distribution of proteins, albeit several unique proteins are predicted from different pipelines. This exercise suggests that some gene(s) could get overlooked when using a single annotation protocol. Therefore, multiple annotations i.e., RAST, GLIMMER, and xBASE were used for comparative studies, whereas RAST annotations were used for genome-based studies for uniformity.

### Phylogenetic analysis of the *Myxococcus* clade

In genus *Myxococcus*, more than eight species had been reported many of which have been taxonomically reclassified in the absence of respective type strains [44, 45]. Presently this genus consists of *M. xanthus*, *M. virescens*, *M. flavescens*, *M. stipitatus* and *M. macrosporus*, which differ in the morphology of their vegetative cells and fruiting body, along with pigment formation during swarm growth [45]. All species exhibit typical long and rod-shaped morphology during vegetative states with varied cell sizes [45]. During fruiting body formation, these bacteria display diverse and distinct morphology [20, 46]. Given their close relationship and overlapping morphological features, the taxonomic placement of *Myxococcus* strains is difficult. For instance, in literature *M. macrosporus* has been referred to as *Coralloccoccus macrosporus* but is regarded as a species of the *Myxococcus* genera [47]. Here we discuss the taxonomic position of *M. hansupus* based on 16S rRNA, housekeeping genes, and genome-genome distance based phylogeny. The resulting tree from 16S rRNA sequences was not able to resolve all species of the *Myxococcus* genus (S1 Fig). The sequence similarities within all *Myxococcus* spp. 16S rRNA





**Fig 4. Phylogeny based on housekeeping proteins.** Twenty-eight concatenated housekeeping proteins were used to generate ML based phylogenetic tree using MEGA 6.06 [model: JTT matrix; bootstrap: 100]. *Corallocooccus coralloides* DSM 2259, *Cystobacter fuscus* DSM 2262, *Anaeromyxobacter dehalogenans* 2CP-C, *Sorangium cellulosum* Soce56, and *Bdellovibrio exovorius* JSS were used as outgroup species in this study. Bootstrap values corresponding to the tree nodes are provided.

doi:10.1371/journal.pone.0148593.g004

sequences were more than 96%. *Mh* shows maximum similarity with *Mf* (99.45%) followed by *Mx* (99.24%) and then *Ms* (98.28%) suggesting its close relationship to *Mf*. The tree correctly grouped *M. xanthus* strains i.e. *MxDK1622*, *MxDZ2* and *MxDZF1* strains, confirming their closeness with each other. Some irregularities in the taxonomic tree include the positions of *M. flavescens* NBRC 100081 and *M. flavescens* NBRC100077 similar to what has been reported previously [45].

In spite of its popularity, 16S rRNA is not a credible marker for taxonomic placement below the genus level [48], therefore housekeeping gene analysis was performed to validate the taxonomic relationship among *Myxococcus* genus. The phylogenetic analysis of 28 housekeeping genes [34] (Fig 4) of the *Myxococcus* clade and five outgroups, reveals a similar tree topology as obtained using 16S rRNA and supports the assertion that *Mh* is closely related to *Mf* followed by other *Mx* species. *MxDK1622*, *MxDZ2*, and *MxDZF1* were placed together, similar to the 16S rRNA based tree. We have also estimated DNA-DNA hybridization (DDH) values between genus *Myxococcus* members using GGDC (Genome-To-Genome Distance Calculator) server [49] which uses GBDP strategy (Genome Blast Distance Phylogeny) (S1 Table). *Mh* genome shares lowest intergenomic distance (highest DDH value) with *Mf* genome. The maximum DDH value of 62.1% with *Mf* further suggests that *Mh* is a novel species within the genus *Myxococcus* as for two organisms to belong to the same species, DDH value should be greater than 70% [49].

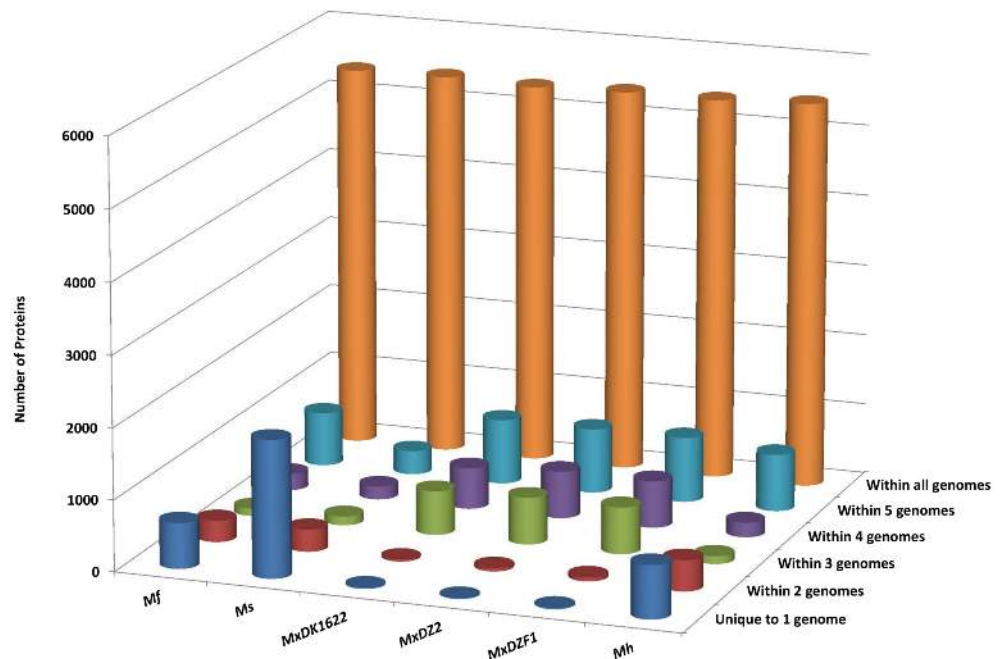
### Pan Proteome analysis: Core, Dispensable, and Unique proteome

The Pan Proteome is defined as the sum total of protein content associated with more than two species; and consists of the Core Proteome, Dispensable Proteome and Unique Proteome [50,

51]. The total proteome of six *Myxococcus* genomes consists of 46,392 proteins with 7,901 orthologous protein clusters (S2 Table), where a cluster signifies one representative from each genome. The percentages of proteins from different members in the clusters are *Mf*: 82.09%, *Ms*: 66.44%, *MxDK1622*: 94.90%, *Mh*: 81.39%, *MxDZF1*: 94.27% and *MxDZ2*: 94.33%. Among these, 4,693 clusters are found to be conserved in all the genomes and define the core proteome for the *Myxococcus* clade. The core proteome accounts for 56.6–63% of the total protein content in each genome and consist of genes that are responsible for essential biological functions such as homeostasis, housekeeping functions and maintaining morphological, developmental and physiological features of the organism. The function profile analysis of the core proteome depicts that 5.4% of the proteins are involved in signal transduction while 45% of the core proteome is involved in housekeeping functions such as cell wall/membrane biogenesis (M), amino acid transport (E), translation and ribosome biogenesis (J), post-translational modifications (O), energy production (C), lipid transport (I), replication (L), carbohydrate transport (G), secondary metabolism biosynthesis (Q) and cell motility (N). Fifteen percent of the proteins were assigned to COG's general function (R) category while the remaining 35% could not be attributed to any known function. The proteins sharing orthology within two or more genomes, but not in all genomes under study, are defined as the dispensable proteome. The dispensable proteome varies from 9.85% (in *Ms*) to 33% (*Mx* strains) among the genomes, a majority of which is likely involved in species-specific functions. The dispensable proteome consists of genes that allow the organism to sustain its species level diversity and participate in the regulation of accessory functions [50]. The analysis reveals that *M. stipitatus* show the minimum orthology protein pairs with other species followed by *Mh* and *Mf*.

Homology studies among the genomes provide insights into the extent of duplicated genes, thereby explaining an important factor of genome expansion. Homologous genes among all the genomes and unique genes in each of the genome were identified in this study. There are 46,392 proteins encoded by all six genomes out of which 32,415 proteins (69.87% of total proteins) have homologs in all genomes, which accounts for 5,453 proteins in *Mh*; 5,395 in *MxDZF1*; 5,401 in *MxDZ2*; 5,367 in *MxDK1622*; 5,406 in *Ms* and 5,393 proteins in *Mf*; representing 70.33, 70.06, 70.24, 71.33, 65.19 and 72.55% of proteins from each genome (Fig 5). The remaining proteins are either restricted to a single genome or present in two or more genomes (Table 3). An all-to-all protein content comparison matrix reveals that *Mh* shares 82.7% genes of *Mf* and 83.36% of *MxDK1622* while only 76.82% genes of *Ms* are mapped to *Mh* (S3 Table). Likewise, *Mf*, *MxDK1622*, and *Ms* share 85.48%, 84.77% and 71.92% genes of *Mh* respectively. *MxDK1622*, *MxDZ2*, and *MxDZF1* are quite similar to each other, with only 0.5–1.0% difference in their protein content. This suggests that the genomes herein share ~80% of their protein content while diversity and uniqueness in each genome are achieved by the remaining 20% of the genes. Complete chromosomes of *M. hansupus*, *M. fulvus* HW-1, *M. stipitatus* DSM 14675 and *M. xanthus* DK1622 were aligned with each other and syntenic plots for all combinations of genomes were generated (S2 Fig). Blue and red dots represent putative homologous regions in positive and negative DNA direction between two genomes as identified by sequence similarity. These plots revealed large identical syntenic blocks suggesting relative closeness between the genus *Myxococcus* genomes. We also identified various insertions and translocations within these genomes.

Unique proteins were also identified using BLAST analysis. These proteins are present only in one genome with no homologs in other genomes. The number of unique proteins varies from 12 to 1,929 in genus *Myxococcus* members (Fig 5). This account for 0.16% unique proteins in *MxDZ2*, 0.23% in *MxDZF1*, 0.25% in *MxDK1622*, 8.65% in *Mf*, 9.63% in *Mh* and 23.26% in *Ms*. The large number of unique proteins, mostly annotated as hypothetical proteins with unknown functions, is suggestive of high genomic diversity within the same genus.



**Fig 5. Comparative representation of homologous protein distribution within the genomes.** Proteome dataset from each genome was subjected to BLASTp to identify homologous proteins between the genomes with an E-value cutoff of  $1e-5$ , query coverage of 50% and identity of 35%. Protein distribution between different combinations of genomes was identified and is represented as a 3D-graph. X, Y and Z-axis respectively denote genome name, the number of proteins and the genome combination.

doi:10.1371/journal.pone.0148593.g005

## Protein clustering

We have clustered each proteome dataset, to compare the homologous proteins within each genome. Our clustering analysis suggests that these six genomes have 590–660 protein clusters sharing on an average of 2424 proteins in each genome which may be represented as multi-copy or duplicated proteins (S4 Table). Each cluster contains between 2 to 431 proteins. The remaining, 5308 proteins on average are singletons and have no homologs within the genome. Our analysis suggests that on average 31.33% of the proteins in each genome are present in multiple copies in the *Myxococcus* genomes. Among these duplicated proteins the maximum representation was from response regulators, protein kinases, ABC transporters, long-chain fatty acid CoA ligase, short-chain dehydrogenase and LysR family transcriptional regulator proteins.

We also performed Pfam domain and clan-based clustering for the six *Myxococcus* proteomes along with rest of the Myxobacteria (*Sorangium*, *Cystobacter*, *Chondromyces*, *Plesiocystis*, *Stigmatella*, *Corallococcus*, *Haliangium*, and *Anaeromyxobacter*) and representative proteomes of non-Myxococcales  $\delta$ -proteobacteria (S5 Table). We found that several Pfam clans such as protein kinase domain [CL0016], PP-binding (CL0314), PKinase (CL0016), CoA-acyltrans (CL0149), Peptidase\_PA (CL0124), GroES (CL0296), AB\_hydrolase (CL0028), Thiolase (CL0046), CheY (CL0304), AMP-binding\_C (CL0531), HTH (CL0123) etc. are overrepresented in order Myxococcales members by more than 200% as compared to non-Myxococcales  $\delta$ -proteobacteria. Besides this, many Pfam clans such as EGF (CL0001), Trefoil (CL0066), gCrystallin (CL0333), Aerolisin\_ETX (CL0345), Hydrophilin (CL0385), Viral\_Gag (CL0148), zf-FYVE-PHD (CL0390), HMG-box (CL0114), PLAT (CL0321), EsxAB (CL0352), Frag1-like

Table 3. A binary map depicting cluster of homologous proteins among all genomes.

<i>M. fulvus</i>	<i>M. stipitatus</i>	<i>M. xanthus</i> DK1622	<i>M. xanthus</i> DZ2	<i>M. xanthus</i> DZF1	<i>M. hansupus</i>	Presence in genomes	Total proteins	Total proteins %
a	a	a	a	a	P	1	747	1.61
a	a	a	a	P	a	1	18	0.04
a	a	a	P	a	a	1	12	0.03
a	a	P	a	a	a	1	19	0.04
a	P	a	a	a	a	1	1929	4.16
P	a	a	a	a	a	1	643	1.39
a	a	a	a	P	P	2	3	0.01
a	a	P	a	a	P	2	5	0.01
a	P	a	a	a	P	2	430	0.93
P	a	a	a	a	P	2	453	0.98
a	a	a	P	P	a	2	77	0.17
a	a	P	a	P	a	2	30	0.07
P	a	a	a	P	a	2	1	0.00
a	a	P	P	a	a	2	16	0.03
a	P	a	P	a	a	2	1	0.00
P	a	a	P	a	a	2	1	0.00
a	P	P	a	a	a	2	2	0.00
P	a	P	a	a	a	2	4	0.01
P	P	a	a	a	a	2	157	0.34
a	a	a	P	P	P	3	8	0.02
a	a	P	a	P	P	3	1	0.00
P	a	a	a	P	P	3	4	0.01
a	a	P	P	a	P	3	5	0.01
a	P	P	a	a	P	3	8	0.02
P	a	P	a	a	P	3	3	0.01
P	P	a	a	a	P	3	295	0.64
a	a	P	P	P	a	3	1869	4.03
a	P	a	P	P	a	3	18	0.04
P	a	a	P	P	a	3	18	0.04
a	P	P	a	P	a	3	4	0.01
P	a	P	a	P	a	3	8	0.02
a	P	P	P	a	a	3	12	0.03
P	a	P	P	a	a	3	9	0.02
P	P	P	a	a	a	3	4	0.01
a	a	P	P	P	P	4	830	1.79
a	P	a	P	P	P	4	1	0.00
P	a	a	P	P	P	4	17	0.04
a	P	P	a	P	P	4	2	0.00
P	a	P	a	P	P	4	8	0.02
P	P	a	a	P	P	4	2	0.00
a	P	P	P	a	P	4	6	0.01
P	a	P	P	a	P	4	1	0.00
P	P	a	P	a	P	4	4	0.01
P	P	P	a	a	P	4	11	0.02
a	P	P	P	P	a	4	653	1.41
P	a	P	P	P	a	4	976	2.10

(Continued)

Table 3. (Continued)

<i>M. fulvus</i>	<i>M. stipitatus</i>	<i>M. xanthus</i> DK1622	<i>M. xanthus</i> DZ2	<i>M. xanthus</i> DZF1	<i>M. hansupus</i>	Presence in genomes	Total proteins	Total proteins %
P	P	a	P	P	a	4	1	0.00
P	P	P	P	a	a	4	15	0.03
a	P	P	P	P	P	5	870	1.88
P	a	P	P	P	P	5	3148	6.79
P	P	a	P	P	P	5	10	0.02
P	P	P	a	P	P	5	24	0.05
P	P	P	P	a	P	5	13	0.03
P	P	P	P	P	a	5	571	1.23
P	P	P	P	P	P	6	32415	69.87

The map depicts the count of homologous proteins along with their presence [denoted as ‘P’] and absence [denoted as ‘a’] in the genome combinations. The numbers in the last column represent the fraction of the proteins from all genomes. Proteins present in various combinations such as one (unique), two, three, four, five and all genomes are shown.

doi:10.1371/journal.pone.0148593.t003

(CL0412), Hexosaminidase (CL0546) etc. are particularly present in genus *Myxococcus* members and not in the non-Myxococcales  $\delta$ -proteobacteria. The presence of overrepresented and unique Pfam families in *Myxococcus* genomes as compared to other non-Myxococcales  $\delta$ -proteobacterial genomes is suggestive of the nature of genome expansion and could probably help these organisms to adapt to diverse habitats and in leading a complex life cycle. Such adaptability could have been achieved by gain, loss or duplication of gene/protein content [19, 52]. Our results are in accordance with reports that attribute gene duplication as one of the main driving force behind genome expansion in *Myxococcus* genomes [42].

### Conclusion

The current study reports the complete 9.5 Mbp genome of a novel Myxobacteria, *M. hansupus* and its comparative analysis with five previously available *Myxococcus* genomes. 16S rRNA, housekeeping genes phylogeny, and genome-genome distance suggest this organism is a novel species of the genus *Myxococcus*. Overall protein similarity among six *Myxococcus* genomes, which include four different species and three strains of *M. xanthus*, help define the core, dispensable and unique proteomes for genus *Myxococcus*. Orthology analysis revealed ~60% of the proteins as the core proteome whereas homology studies identified the presence of ~70% of the total proteome in these closely related genus *Myxococcus* members. The wide genome diversity at species level within genus *Myxococcus* is revealed by the presence of large number of unique proteins, e.g. as high as 1,929 unique proteins in *M. stipitatus* genome. Protein sequence clustering reveals that 31% of the total protein content is present in multiple copies with a majority of the proteins functioning as response regulators, kinases and ABC transporters. The presence of several overrepresented Pfam clans and their constituting families helps in identifying the genome expansion in *Myxococcus* genomes as compared to other non-Myxococcales  $\delta$ -proteobacteria genomes.

### Supporting Information

**S1 Fig. Phylogenetic analysis of genus *Myxococcus* 16S rRNA.** MEGA 6.06 was used to generate a maximum likelihood tree [model: Tamura 3-param; bootstrap: 1000]. Different leaf colors were used in the tree to demarcate species; *M. xanthus*: navy blue, *M. fulvus*: dark teal, *M.*

*stipitatus*: yellow, *M. virescens*: green, *M. macrosporus*: dark brown and *M. flavescens*: light brown. Black circle represents complete genomes and red semicircle represents the draft genomes. Bootstrap values are provided corresponding to the tree nodes. *Coralloccoccus coralloides*, *Cystobacter fuscus*, *Anaeromyxobacter dehalogenans*, *Sorangium cellulosum*, and *Bdellovibrio exovorus* were used as outgroup species in this study.

(EPS)

**S2 Fig. Syntenic dot plot between genus *Myxococcus* complete genomes.** Complete genomes of *M. hansupus*, *M. fulvus* HW-1, *M. stipitatus* DSM 14675 and *M. xanthus* DK1622 were used in this study. Blue and red dots represent putative homologous regions between two genomes in the positive and negative directions respectively as identified by sequence similarity. Panel A, B and C represent the dot plots for *M. hansupus* aligned against *M. xanthus*, *M. fulvus* and *M. stipitatus* respectively. Panel D and E represent the dot plots for *M. fulvus* aligned against *M. stipitatus* and *M. xanthus*. Panel F represents the dot plot analysis of *M. xanthus* aligned against *M. stipitatus*.

(EPS)

**S1 Table. Genome-to-genome distance between *M. hansupus* and its neighbors.** The genome-to-genome distance between *M. hansupus* and other myxobacterial genomes (*Coralloccoccus coralloides* DSM 2259, *Cystobacter fuscus* DSM 2262, *Anaeromyxobacter dehalogenans* 2CP-1, and *Sorangium cellulosum* Soce56) and *Bdellovibrio bacteriovorus* W using GGDC (version 2.0 and Formula 2) is shown here. Red-yellow-green shading depicts decreasing closeness based on DDH values.

(XLS)

**S2 Table. Matrix of orthologous protein sets for six *Myxococcus* genomes.** The matrix depicts the count of orthologous proteins [column J] along with their presence [denoted as P in white shade] and absence [shaded black] in all possible genome combinations [column I]. Orthologous proteins present in two, three, four, five and all genomes are shaded gray, light green, dark purple, blue and dark green respectively.

(XLS)

**S3 Table. All-to-all protein content comparison matrix.** All proteins were mapped between two genomes and their mapping percentage to each genome is represented here with high to low (red-yellow-green) shading order. The matrix should be read as % proteins of [row] genome mapped against [column] genome.

(XLS)

**S4 Table. Protein clustering analysis between *Myxococcus* genomes.** BLASTp results were filtered on the basis of cut-off values [E-value:  $1e^{-5}$ , query coverage: 50% and identity: 35%] and protein homologs were clustered. Singleton proteins, proteins present in clusters and numbers of clusters are shown in column D, E and G.

(XLS)

**S5 Table. Pfam clans based clustering in six *Myxococcus* genomes and comparative distribution with rest of the myxobacteria and non-Myxococcales  $\delta$ -proteobacterial members.** Apart from six *Myxococcus* genomes [column C-H], rest of the myxobacteria [column K] include *Sorangium*, *Cystobacter*, *Chondromyces*, *Plesiocystis*, *Stigmatella*, *Coralloccoccus*, *Haliangium*, and *Anaeromyxobacter*. Non-Myxococcales  $\delta$ -proteobacteria [column L] include *Bacteriovorax marinus* SJ, *Bdellovibrio bacteriovorus* HD100, *Bilophila wadsworthia*, *Deferrisoma camini*, *Desulfarculus baarsii* DSM 2075, *Desulfatibacillum alkenivorans* AK 01, *Desulfatiglans anilini*, *Desulfatirhabdium butyrativorans*, *Desulfobacca acetoxidans* DSM 11109,

*Desulfobacter curvatus*, *Desulfobacterium autotrophicum* HRM2, *Desulfobacula toluolica* Tol2, *Desulfobulbaceae* bacterium BRH c16a, *Desulfobulbus propionicus* DSM 2032, *Desulfocapsa sulfexigens* DSM 10523, *Desulfococcus oleovorans* Hxd3, *Desulfocurvus vexinensis*, *Desulfohalobium retbaense* DSM 5692, *Desulfomicrobium baculatum* DSM 4028, *Desulfomonile tiedjei* DSM 6799, *Desulfonatronum thioautotrophicum*, *Desulforegula conservatrix*, *Desulfotalea psychrophila* LSv54, *Desulfotignum balticum*, *Desulfovermiculus halophilus*, *Desulfovibrio hydrothermalis* AM13, *Desulfurella acetivorans*, *Desulfurivibrio alkaliphilus* AHT2, *Desulfuromonas acetoxidans*, *Geoalkalibacter ferrihydriticus*, *Geobacter sulfurreducens* KN400, *Geopsychrobacter electrophilus*, *Hipaea maritima* DSM 10411, *Lawsonia intracellularis* N343, *Pelobacter carbinolicus* DSM 2380, *Syntrophobacter fumaroxidans* MPOB, *Syntrophorhabdus aromaticivorans* and *Syntrophus aciditrophicus* SB. Pfam clan number and clan name are shown in column A and B. The numbers of proteins per clan in six *Myxococcus* genomes are represented [column C-H] with high to low (red-yellow-green) shading order. A similar shading is used for the numbers of proteins per clan in average *Myxococcus* [column J], average rest-Myxobacteria [column K] and average-non-Myxococcales  $\delta$ -proteobacteria [column L]. In column M, % increase/decrease of numbers of proteins per clan in *Myxococcus* is depicted as compared to  $\delta$ -proteobacteria. (XLS)

## Acknowledgments

This work is supported by a project "Expansion and modernization of Microbial Type Culture Collection and Gene Bank (MTCC) jointly supported by Council of Scientific and Industrial Research (CSIR) Grant No. BSC0402 and Department of Biotechnology (DBT) Govt. of India Grant No. BT/PR7368/INF/22/177/2012". GS acknowledges CSIR for research fellowship. TN is supported by the CSIR network program GENESIS-CSIR-BSC-0121. We also thank Dr. Ramya TNC for providing us workspace and guidance in myxobacterial culturing. We thank the Genome Quebec Innovation Center, McGill University, Canada for help in obtaining Pac-Bio sequencing data. We also thank CSIR-IMTECH Electron Microscopy facility for SEM data.

## Author Contributions

Conceived and designed the experiments: SS. Performed the experiments: GS. Analyzed the data: GS TN SS. Contributed reagents/materials/analysis tools: GS TN. Wrote the paper: GS TN SS. Coded PERL scripts used in the analysis: TN.

## References

1. Roland T. Contributions from the Cryptogamic Laboratory of Harvard University. LVI. Notes on the Myxobacteriaceae. *Botanical Gazette*. 1904; 37(6):405–16.
2. Velicer GJ, Vos M. Sociobiology of the myxobacteria. *Annual review of microbiology*. 2009; 63:599–623. Epub 2009/07/07. doi: [10.1146/annurev.micro.091208.073158](https://doi.org/10.1146/annurev.micro.091208.073158) PMID: [19575567](https://pubmed.ncbi.nlm.nih.gov/19575567/).
3. Sanford RA, Cole JR, Tiedje JM. Characterization and Description of *Anaeromyxobacter dehalogenans* gen. nov., sp. nov., an Aryl-Halo-respiring Facultative Anaerobic Myxobacterium. *Applied and Environmental Microbiology*. 2002; 68(2):893–900. doi: [10.1128/aem.68.2.893-900.2002](https://doi.org/10.1128/aem.68.2.893-900.2002) PMID: [11823233](https://pubmed.ncbi.nlm.nih.gov/11823233/)
4. Mauriello EM, Mignot T, Yang Z, Zusman DR. Gliding motility revisited: how do the myxobacteria move without flagella? *Microbiology and molecular biology reviews: MMBR*. 2010; 74(2):229–49. Epub 2010/05/29. doi: [10.1128/MMBR.00043-09](https://doi.org/10.1128/MMBR.00043-09) PMID: [20508248](https://pubmed.ncbi.nlm.nih.gov/20508248/); PubMed Central PMCID: [PMC2884410](https://pubmed.ncbi.nlm.nih.gov/PMC2884410/).
5. Huntley S, Hamann N, Wegener-Feldbrugge S, Treuner-Lange A, Kube M, Reinhardt R, et al. Comparative genomic analysis of fruiting body formation in Myxococcales. *Molecular biology and evolution*. 2011; 28(2):1083–97. Epub 2010/11/03. doi: [10.1093/molbev/msq292](https://doi.org/10.1093/molbev/msq292) PMID: [21037205](https://pubmed.ncbi.nlm.nih.gov/21037205/).

6. Wu Y, Jiang Y, Kaiser AD, Alber M. Self-organization in bacterial swarming: lessons from myxobacteria. *Physical biology*. 2011; 8(5):055003. Epub 2011/08/13. doi: [10.1088/1478-3975/8/5/055003](https://doi.org/10.1088/1478-3975/8/5/055003) PMID: [21832807](https://pubmed.ncbi.nlm.nih.gov/21832807/).
7. Berleman JE, Kirby JR. Deciphering the hunting strategy of a bacterial wolfpack. *FEMS microbiology reviews*. 2009; 33(5):942–57. Epub 2009/06/13. doi: [10.1111/j.1574-6976.2009.00185.x](https://doi.org/10.1111/j.1574-6976.2009.00185.x) PMID: [19519767](https://pubmed.ncbi.nlm.nih.gov/19519767/); PubMed Central PMCID: PMC2774760.
8. Wolgemuth C, Hoiczky E, Kaiser D, Oster G. How myxobacteria glide. *Current biology: CB*. 2002; 12(5):369–77. Epub 2002/03/08. PMID: [11882287](https://pubmed.ncbi.nlm.nih.gov/11882287/).
9. Nan B, Zusman DR. Uncovering the mystery of gliding motility in the myxobacteria. *Annual review of genetics*. 2011; 45:21–39. Epub 2011/09/14. doi: [10.1146/annurev-genet-110410-132547](https://doi.org/10.1146/annurev-genet-110410-132547) PMID: [21910630](https://pubmed.ncbi.nlm.nih.gov/21910630/); PubMed Central PMCID: PMC3397683.
10. Kaiser D, Robinson M, Kroos L. Myxobacteria, polarity, and multicellular morphogenesis. *Cold Spring Harbor perspectives in biology*. 2010; 2(8):a000380. Epub 2010/07/09. doi: [10.1101/cshperspect.a000380](https://doi.org/10.1101/cshperspect.a000380) PMID: [20610548](https://pubmed.ncbi.nlm.nih.gov/20610548/); PubMed Central PMCID: PMC2908774.
11. Shimkets LJ. Social and developmental biology of the myxobacteria. *Microbiological reviews*. 1990; 54(4):473–501. Epub 1990/12/01. PMID: [1708086](https://pubmed.ncbi.nlm.nih.gov/1708086/); PubMed Central PMCID: PMC372790.
12. Han K, Li ZF, Peng R, Zhu LP, Zhou T, Wang LG, et al. Extraordinary expansion of a *Sorangium cellulosum* genome from an alkaline milieu. *Scientific reports*. 2013; 3:2101. Epub 2013/07/03. doi: [10.1038/srep02101](https://doi.org/10.1038/srep02101) PMID: [23812535](https://pubmed.ncbi.nlm.nih.gov/23812535/); PubMed Central PMCID: PMC3696898.
13. Schneiker S, Perlova O, Kaiser O, Gerth K, Alici A, Altmeyer MO, et al. Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nature biotechnology*. 2007; 25(11):1281–9. Epub 2007/10/30. doi: [10.1038/nbt1354](https://doi.org/10.1038/nbt1354) PMID: [17965706](https://pubmed.ncbi.nlm.nih.gov/17965706/).
14. Wang Y, Yang JK, Lee OO, Li TG, Al-Suwailem A, Danchin A, et al. Bacterial Niche-Specific Genome Expansion Is Coupled with Highly Frequent Gene Disruptions in Deep-Sea Sediments. *PLoS One*. 2011; 6(12):e29149. doi: [10.1371/journal.pone.0029149](https://doi.org/10.1371/journal.pone.0029149) PMID: [22216192](https://pubmed.ncbi.nlm.nih.gov/22216192/)
15. Goldman B, Bhat S, Shimkets LJ. Genome evolution and the emergence of fruiting body development in *Myxococcus xanthus*. *PLoS One*. 2007; 2(12):e1329. Epub 2007/12/27. doi: [10.1371/journal.pone.0001329](https://doi.org/10.1371/journal.pone.0001329) PMID: [18159227](https://pubmed.ncbi.nlm.nih.gov/18159227/); PubMed Central PMCID: PMC2129111.
16. Huntley S, Kneip S, Treuner-Lange A, Sogaard-Andersen L. Complete genome sequence of *Myxococcus stipitatus* strain DSM 14675, a fruiting myxobacterium. *Genome announcements*. 2013; 1(2):e0010013. Epub 2013/03/22. doi: [10.1128/genomeA.00100-13](https://doi.org/10.1128/genomeA.00100-13) PMID: [23516218](https://pubmed.ncbi.nlm.nih.gov/23516218/); PubMed Central PMCID: PMC3622980.
17. Muller S, Willett JW, Bahr SM, Scott JC, Wilson JM, Darnell CL, et al. Draft Genome of a Type 4 Pilus Defective *Myxococcus xanthus* Strain, DZF1. *Genome announcements*. 2013; 1(3). Epub 2013/06/22. doi: [10.1128/genomeA.00392-13](https://doi.org/10.1128/genomeA.00392-13) PMID: [23788552](https://pubmed.ncbi.nlm.nih.gov/23788552/); PubMed Central PMCID: PMC3707601.
18. Muller S, Willett JW, Bahr SM, Darnell CL, Hummels KR, Dong CK, et al. Draft Genome Sequence of *Myxococcus xanthus* Wild-Type Strain DZ2, a Model Organism for Predation and Development. *Genome announcements*. 2013; 1(3). Epub 2013/05/11. doi: [10.1128/genomeA.00217-13](https://doi.org/10.1128/genomeA.00217-13) PMID: [23661486](https://pubmed.ncbi.nlm.nih.gov/23661486/); PubMed Central PMCID: PMC3650445.
19. Brattie MS, Johansen J, Sherman BT, Huang da W, Lempicki RA, Drablos F. Gene duplications in prokaryotes can be associated with environmental adaptation. *BMC genomics*. 2010; 11:588. Epub 2010/10/22. doi: [10.1186/1471-2164-11-588](https://doi.org/10.1186/1471-2164-11-588) PMID: [20961426](https://pubmed.ncbi.nlm.nih.gov/20961426/); PubMed Central PMCID: PMC3091735.
20. Whitworth DE. *Myxobacteria: Multicellularity and Differentiation: American Society of Microbiology*; 2008.
21. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Meth*. 2013; 10(6):563–9. doi: [10.1038/nmeth.2474](https://doi.org/10.1038/nmeth.2474)
22. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. *BMC genomics*. 2008; 9:75. Epub 2008/02/12. doi: [10.1186/1471-2164-9-75](https://doi.org/10.1186/1471-2164-9-75) PMID: [18261238](https://pubmed.ncbi.nlm.nih.gov/18261238/); PubMed Central PMCID: PMC2265698.
23. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic acids research*. 2007; 35(9):3100–8. Epub 2007/04/25. doi: [10.1093/nar/gkm160](https://doi.org/10.1093/nar/gkm160) PMID: [17452365](https://pubmed.ncbi.nlm.nih.gov/17452365/); PubMed Central PMCID: PMC1888812.
24. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic acids research*. 1997; 25(5):955–64. Epub 1997/03/01. PMID: [9023104](https://pubmed.ncbi.nlm.nih.gov/9023104/); PubMed Central PMCID: PMC146525.
25. Goldman BS, Nierman WC, Kaiser D, Slater SC, Durkin AS, Eisen JA, et al. Evolution of sensory complexity recorded in a myxobacterial genome. *Proceedings of the National Academy of Sciences of the*



- United States of America. 2006; 103(41):15200–5. doi: [10.1073/pnas.0607335103](https://doi.org/10.1073/pnas.0607335103) PMID: [17015832](https://pubmed.ncbi.nlm.nih.gov/17015832/); PubMed Central PMCID: PMC1622800.
26. Li ZF, Li X, Liu H, Liu X, Han K, Wu ZH, et al. Genome sequence of the halotolerant marine bacterium *Myxococcus fulvus* HW-1. *Journal of bacteriology*. 2011; 193(18):5015–6. Epub 2011/08/27. doi: [10.1128/jb.05516-11](https://doi.org/10.1128/jb.05516-11) PMID: [21868801](https://pubmed.ncbi.nlm.nih.gov/21868801/); PubMed Central PMCID: PMC3165639.
  27. Gao F, Zhang CT. OriC: a database of oriC regions in bacterial genomes. *Bioinformatics*. 2007; 23(14):1866–7. Epub 2007/05/15. doi: [10.1093/bioinformatics/btm255](https://doi.org/10.1093/bioinformatics/btm255) PMID: [17496319](https://pubmed.ncbi.nlm.nih.gov/17496319/).
  28. Gao F, Zhang CT. Ori-Finder: a web-based system for finding oriCs in unannotated bacterial genomes. *BMC bioinformatics*. 2008; 9:79. Epub 2008/02/02. doi: [10.1186/1471-2105-9-79](https://doi.org/10.1186/1471-2105-9-79) PMID: [18237442](https://pubmed.ncbi.nlm.nih.gov/18237442/); PubMed Central PMCID: PMC3165639.
  29. Aggarwal G, Ramaswamy R. Ab initio gene identification: prokaryote genome annotation with GeneScan and GLIMMER. *Journal of biosciences*. 2002; 27(1 Suppl 1):7–14. Epub 2002/04/03. PMID: [11927773](https://pubmed.ncbi.nlm.nih.gov/11927773/).
  30. Chaudhuri RR, Pallen MJ. xBASE, a collection of online databases for bacterial comparative genomics. *Nucleic acids research*. 2006; 34(Database issue):D335–7. Epub 2005/12/31. doi: [10.1093/nar/gkj140](https://doi.org/10.1093/nar/gkj140) PMID: [16381881](https://pubmed.ncbi.nlm.nih.gov/16381881/); PubMed Central PMCID: PMC1347502.
  31. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*. 1999; 41:95–8. citeulike-article-id:691774.
  32. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution*. 2013; 30(12):2725–9. Epub 2013/10/18. doi: [10.1093/molbev/mst197](https://doi.org/10.1093/molbev/mst197) PMID: [24132122](https://pubmed.ncbi.nlm.nih.gov/24132122/); PubMed Central PMCID: PMC3840312.
  33. Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic acids research*. 2011; 39(Web Server issue):W475–8. Epub 2011/04/08. doi: [10.1093/nar/gkr201](https://doi.org/10.1093/nar/gkr201) PMID: [21470960](https://pubmed.ncbi.nlm.nih.gov/21470960/); PubMed Central PMCID: PMC3125724.
  34. Wu M, Eisen J. A simple, fast, and accurate method of phylogenomic inference. *Genome Biology*. 2008; 9(10):R151. PMID: [18851752](https://pubmed.ncbi.nlm.nih.gov/18851752/)
  35. Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-) orthologs in large-scale analysis. *BMC bioinformatics*. 2011; 12:124. Epub 2011/04/30. doi: [10.1186/1471-2105-12-124](https://doi.org/10.1186/1471-2105-12-124) PMID: [21526987](https://pubmed.ncbi.nlm.nih.gov/21526987/); PubMed Central PMCID: PMC3114741.
  36. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of molecular biology*. 1990; 215(3):403–10. Epub 1990/10/05. doi: [10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: [2231712](https://pubmed.ncbi.nlm.nih.gov/2231712/).
  37. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. *Genome Biol*. 2004; 5(2):R12. Epub 2004/02/05. PMID: [14759262](https://pubmed.ncbi.nlm.nih.gov/14759262/); PubMed Central PMCID: PMC395750.
  38. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. *Nucleic acids research*. 2014; 42(1):D222–30. Epub 2013/11/30. doi: [10.1093/nar/gkt1223](https://doi.org/10.1093/nar/gkt1223) PMID: [24288371](https://pubmed.ncbi.nlm.nih.gov/24288371/).
  39. Eddy SR. Accelerated Profile HMM Searches. *PLoS computational biology*. 2011; 7(10):e1002195. Epub 2011/11/01. doi: [10.1371/journal.pcbi.1002195](https://doi.org/10.1371/journal.pcbi.1002195) PMID: [22039361](https://pubmed.ncbi.nlm.nih.gov/22039361/); PubMed Central PMCID: PMC3197634.
  40. Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic acids research*. 2015; 43(Database issue):D298–9. Epub 2014/11/08. doi: [10.1093/nar/gku1046](https://doi.org/10.1093/nar/gku1046) PMID: [25378308](https://pubmed.ncbi.nlm.nih.gov/25378308/); PubMed Central PMCID: PMC3438389.
  41. Zusman DR, Scott AE, Yang Z, Kirby JR. Chemosensory pathways, motility and development in *Myxococcus xanthus*. *Nature reviews Microbiology*. 2007; 5(11):862–72. Epub 2007/10/09. doi: [10.1038/nrmicro1770](https://doi.org/10.1038/nrmicro1770) PMID: [17922045](https://pubmed.ncbi.nlm.nih.gov/17922045/).
  42. Serres MH, Kerr AR, McCormack TJ, Riley M. Evolution by leaps: gene duplication in bacteria. *Biology direct*. 2009; 4:46. Epub 2009/11/26. PMID: [19930658](https://pubmed.ncbi.nlm.nih.gov/19930658/); PubMed Central PMCID: PMC2787491.
  43. Richardson EJ, Watson M. The automatic annotation of bacterial genomes. *Briefings in Bioinformatics*. 2012. doi: [10.1093/bib/bbs007](https://doi.org/10.1093/bib/bbs007)
  44. Lang E, Kroppenstedt RM, Straubler B, Stackebrandt E. Reclassification of *Myxococcus flavescens* Yamanaka et al. 1990VP as a later synonym of *Myxococcus virescens* Thaxter 1892AL. *International journal of systematic and evolutionary microbiology*. 2008; 58(Pt 11):2607–9. Epub 2008/11/06. doi: [10.1099/ijs.0.2008/001859-0](https://doi.org/10.1099/ijs.0.2008/001859-0) PMID: [18984701](https://pubmed.ncbi.nlm.nih.gov/18984701/).
  45. Miyashita M, Sakane T, Suzuki K, Nakagawa Y. 16S rRNA gene and 16S–23S rRNA gene internal transcribed spacer sequences analysis of the genus *Myxococcus*. *FEMS microbiology letters*. 2008; 282(2):241–5. Epub 2008/03/22. doi: [10.1111/j.1574-6968.2008.01127.x](https://doi.org/10.1111/j.1574-6968.2008.01127.x) PMID: [18355284](https://pubmed.ncbi.nlm.nih.gov/18355284/).
  46. Brenner DJ, Staley JT. *Bergey's Manual® of Systematic Bacteriology*. 2nd edn. ed2005.

47. Lang E, Stackebrandt E. Emended descriptions of the genera *Myxococcus* and *Coralloccoccus*, typification of the species *Myxococcus stipitatus* and *Myxococcus macrosporus* and a proposal that they be represented by neotype strains. Request for an Opinion. *International journal of systematic and evolutionary microbiology*. 2009; 59(Pt 8):2122–8. doi: [10.1099/ijs.0.003566-0](https://doi.org/10.1099/ijs.0.003566-0) PMID: [19567579](https://pubmed.ncbi.nlm.nih.gov/19567579/).
48. Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P. Toward automatic reconstruction of a highly resolved tree of life. *Science (New York, NY)*. 2006; 311(5765):1283–7. Epub 2006/03/04. doi: [10.1126/science.1123061](https://doi.org/10.1126/science.1123061) PMID: [16513982](https://pubmed.ncbi.nlm.nih.gov/16513982/).
49. Auch AF, von Jan M, Klenk H-P, Göker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison 2010.
50. Medini D, Donati C, Tettelin H, Massignani V, Rappuoli R. The microbial pan-genome. *Current opinion in genetics & development*. 2005; 15(6):589–94. Epub 2005/09/28. doi: [10.1016/j.gde.2005.09.006](https://doi.org/10.1016/j.gde.2005.09.006) PMID: [16185861](https://pubmed.ncbi.nlm.nih.gov/16185861/).
51. Lapierre P, Gogarten JP. Estimating the size of the bacterial pan-genome. *Trends in genetics: TIG*. 2009; 25(3):107–10. doi: [10.1016/j.tig.2008.12.004](https://doi.org/10.1016/j.tig.2008.12.004) PMID: [19168257](https://pubmed.ncbi.nlm.nih.gov/19168257/).
52. Kondrashov FA. Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proceedings Biological sciences / The Royal Society*. 2012; 279(1749):5048–57. Epub 2012/09/15. doi: [10.1098/rspb.2012.1108](https://doi.org/10.1098/rspb.2012.1108) PMID: [22977152](https://pubmed.ncbi.nlm.nih.gov/22977152/); PubMed Central PMCID: PMC3497230.
53. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC genomics*. 2011; 12:402. Epub 2011/08/10. doi: [10.1186/1471-2164-12-402](https://doi.org/10.1186/1471-2164-12-402) PMID: [21824423](https://pubmed.ncbi.nlm.nih.gov/21824423/); PubMed Central PMCID: PMCPmc3163573.