Draft genome sequence of *Staphylococcus hominis* strain Hudgins isolated from human skin implicates metabolic versatility and several virulence determinants

Shelby Calkins, M.B. Cougar, Colin Jackson, Jordan Zandler, Garrett C. Hudgins, Radwa A. Hanafy, Connie Budd, Donald P. French, Wouter D. Hoff, Noha Youssef

**A B S T R A C T**

*Staphylococcus hominis* is a predominant member of the human skin microbiome. We here report on the genomic analysis of *Staphylococcus hominis* strain Hudgins that was isolated from the wrist area of human skin. The partial genome assembly of *S. hominis* Hudgins consists of 2,211,863 bp of DNA with 2174 protein-coding genes and 90 RNA genes. Based on the genomic analysis of KEGG pathways, the organism is expected to be a versatile heterotroph potentially capable of hydrolyzing the sugars glucose, fructose, mannose, and the amino acids alanine, aspartate, glutamate, glycine, threonine, cysteine, methionine, valine, isoleucine, leucine, lysine, arginine, phenylalanine, tyrosine, and tryptophan for energy production through aerobic respiration, with occasional lactate and acetate fermentation. Evidence for poly-gamma glutamate capsule and type IV Com system pili were identified in the genome. Based on COG analysis, the genome of *S. hominis* Hudgins clusters away from the previously published *S. hominis* genome ZBW5.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

*S. hominis* strain Hudgins was isolated from the surface of human skin in Stillwater, OK as part of the Student Initiated Microbial Discovery (SIMD) project (introduced in [1]). The genus *Staphylococcus* is a phylogenetically and physiologically diverse genus with members ubiquitously found as part of the skin microbiome [2–4]. Infections have been reported in patients with lowered immunity [3,5,6]. Members of *Staphylococcus hominis* were previously isolated from human-associated [7] and animal-associated microbiomes [8], as well as from the environment [9]. Genomic analysis of strains belonging to the *Staphylococcus hominis* can contribute to our understanding of the molecular mechanisms of opportunistic pathogenesis. Such knowledge could potentially help reduce the occurrence and the severity of such infections in the future. Here we report on the draft genomic sequence, and the detailed annotation and analysis of *Staphylococcus hominis* strain Hudgins with an emphasis on its virulence factors.

2. Materials and methods

2.1. Genome sequencing information

2.1.1. Genome project history

The draft assembly and annotation were completed in 2015–2016.

Table 1 shows the genome project information.

2.1.2. Growth conditions and genomic DNA preparation

*S. hominis* Hudgins was isolated from wrist skin on Tryptic soy agar (TSA) and repeatedly streaked (three times) to obtain a pure culture. To have enough biomass for DNA extraction, the strain was grown overnight at 30 °C on TSA plates. Genomic DNA of high sequencing quality was isolated using the MBio PowerSoil® DNA extraction kit according to manufacturer’s instructions. Negative stain TEM micrographs were obtained using the services of the Oklahoma State University Microscopy Lab. Briefly, the sample was placed on a carbon film TEM grid and allowed to incubate for 2 min, after which the excess liquid was wicked off. Phosphotungestic acid (PTA; 2% w/v) was then added to the grid followed by a 45-second incubation. Excess PTA was wicked off and the grid was allowed to dry before it was visualized using JOEL JEM-2100 transmission electron microscope.

http://dx.doi.org/10.1016/j.gdata.2016.10.003

2213-5960/© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
2.1.3. Genome sequencing and assembly

The genome of *S. hominis* Hudgins was sequenced using the Illumina MiSeq platform at the University of Georgia Genomics Facility using 2 × 300 paired end chemistry and an average library insert size of 700 bp. Quality filtered sequence data were assembled with the short read de brujin graph assembly program Velvet [10]. The assembly settings were a kmer value of 101 bp and a minimum contig coverage value of 7×. The genome project is deposited in GOLD (Genomes On-Linewebsite) and as a result a total of 2270 gene models have been deposited in GenBank under the accession MAYR00000000. The version described in this paper is version MAYR0100000.

2.1.4. Genome annotation

Gene models were created using the prokaryotic gene calling software package prodigal [11], and as a result a total of 2270 gene models were predicted with average gene size of 877 bp. Translated protein sequences were functionally annotated using a combination of NCBI Blast hood analysis [14]. The tree was obtained under Kimura 2-parameter multiple sequence alignments of 16S rRNA genes sequences. Multiple alignment analysis were carried out through the Integrated Microbial Genomes Expert Review (IMG-ER) platform.

2.2. Phylogenetic analysis

A maximum likelihood phylogenetic tree was constructed using multiple sequence alignments of 16S rRNA genes sequences. Multiple sequence alignment was conducted in Mega using ClustalW, as were the selection of the best substitution model, and the maximum likelihood analysis [14]. The tree was obtained using Kimura 2-parameter model with evolutionary rate difference among sites (+G, shape = 0.1836). The substitution rate for transitions were 0.172, and for transversions were 0.039. Escherichia coli isolate ECSD9 was used as the outgroup. Bootstrap values, in percent, are based on 100 replicates.

2.3. Comparative genomics

Previous reports of genomic sequences from human skin-associated *Staphylococcus hominis* include strain ZBW5 [7]. We sought to compare the genome of *Staphylococcus hominis* strain Hudgins to several *Staphylococcus* species (including strain ZBW5) as well as *Staphylococcus haemolyticus* genomes (*n* = 16, IMG IDs: 2648501232, 2657245414, 2617271042, 2617271041, 2609460194, 260241636, 2657245129, 642555160, 2548876781, 2582580999, 2648501918, 651285007, 2548877063, 2623620554, 2654587786, 643886039, 2540341024). We used the “Genome clustering” function on the IMG-ER analysis platform to conduct genomic comparisons based on the COG profile. We also used principal component analysis to compare the genomes based on several genomic features including the genome size, the number of genes, the number of transporters identified, the GC content, the number of non-coding bases, the number of genes belonging to COG categories, as well as the number of genes belonging to each COG category. The PCA analysis was conducted using the “princomp” function in the labdsv library of R [15]. The results were visualized using a biplot, where genomes were represented by stars, and genomic features or COG categories used for comparison were represented by arrows. Further comparative genomics were conducted on strain Hudgins genome and five other *S. hominis* genomes (strains C80, ZBW5, RIT-PI-K, etc).
SK119, and VCU122) and included computing the genomic average nucleotide identity (gANI), alignment fraction (AF) [16], and bidirectional best hits using the nucleotide similarity scanner NSimScan [17], as well as gene homology comparisons using Blastn.

3. Results and discussion

3.1. Classification and features

Cells of strain Hudgins appear to be Gram positive, non-motile aerobic cocci that were arranged in tetrads, pairs, as well as singles and clusters (Fig. 1). Colonies on TSA agar were beige in color.

Within the genus Staphylococcus, 52 species are described with validly published names. Strain Hudgins shares 93.7–99.9% 16S rRNA gene identities with other species in the Staphylococcus genus (Table 2). When compared to other Staphylococcus hominis strains with sequenced genomes, strain Hudgins shares 100% 16S rRNA gene similarity with S. hominis strain C80 (Genbank accession number ACM01000000), and 99% similarity with strains H69 (Genbank accession number LVVO000000000), RIT-PI-K [18], ZBW5 [7], and ShAs1, ShAs2, and ShAs3 [8].

Phylogenetic analysis based on the 16S rRNA gene placed Staphylococcus hominis strain CV21 as the closest taxonomic relative of Staphylococcus hominis strain Hudgins (Table 3, and Fig. 2).

3.2. Genome properties

The genome assembly process produced a contig N50 of 1,549,674 bp with a total genome size of 2,211,863 bp. The GC content was 31.31%. There are 96 RNA genes including 15 ribosomal RNA and 60 tRNA genes. Since all rRNA operons identified in the genome were either located at the start or the end of a contig, the organization of the full ribosomal RNA operon could not be deciphered from the existing assembly. However in some of these operons the 5S and 23S rRNA genes were followed by several tRNA genes. Of the 2270 genes detected, 2174 were protein-coding, of which 79.07% had a function prediction, 73.44% represented a COG functional category, and 2.69% were predicted to have a signal peptide. PSORT [19] classified proteins as 51.3% cytoplasmic, 0.015% extracellular, and 25.99% associated with the membrane. Based on the presence of 139 single copy genes [20], the genome is predicted to be 84.5% complete. Genome statistics are shown in Table 4. The distribution of genes into COG functional categories is shown in Table 5.

3.3. Insights from the genome sequence

Genome analysis of S. hominis Hudgins identified a microorganism/bacterium with a typical Gram positive cell wall structure. We identified genes encoding for the biosynthesis of the polar lipid cardiolipin in the genome. Further genomic analysis suggested that strain Hudgins is a metabolically versatile microorganism with the capability to utilize glucose, fructose, and mannose, alanine, asparagine, glutamine, glycine, serine, cysteine, and proline as carbon and energy sources based on the presence of complete catabolic KEGG pathways in the genome. Aerobic respiration capability was evidenced by the presence of a complete
TCA cycle and electron transport chain with P/V/-type ATPase subunits confirming the aerobic nature of the microorganism. Facultative fermentation capability was also identified in the genome where genes for lactate and fermentation were present. Genomic analysis suggested auxotrophy for asparagine, serine, proline, histidine, vitamin B6, biotin, and lipoic acid. However, 113 ABC and 115 secondary transporters were identified in the genome via comparison of the protein-coding genes against the transporter database [21]. Some of the identified transporters could potentially address these auxotrophies. The above results were not based on biochemical tests and are only suggested by the genomic data available.

Genomic analysis also revealed the absence of genes encoding a complete flagellar assembly in line with the electron micrographs of the organism (Fig. 1). Other structures suggested by genomic analysis that could potentially contribute to the virulence of strain Hudgins include the type IV pilus Com operon (comGA-GH-CC-GD-GE-GF) known to exist in Gram positive bacteria [22] in addition to homologues of the Gram positive sortase that would potentially be employed for actin secretion and pilin polymerization [22,23]. The genome also harbors evidence for the biosynthesis of a poly-gamma glutamate capsule where a cap locus was identified in the genome with homologues of the four essential genes capB-capC-capA-capE [24]. The poly-gamma glutamate capsule is likely tethered to the cell wall based on the presence of the anchorage gene capD encoding for a gamma-glutamyl transpeptidase [24]. The halo around strain Hudgins cells in Fig. 1 could potentially be due to the presence of such capsule. Several other virulence factors are possible as well since comparison of the Hudgins genome against the virulence factor database [25] identified 475 virulence factor hits (21.8% of the protein-coding genes).

3.4. Insights from comparative genomics

We compared the genome of Staphylococcus hominis strain Hudgins to several Staphylococcus hominis (including the previously sequenced Staphylococcus hominis strain ZBW5) as well as Staphylococcus haemolyticus genomes. Based on their COG profile, the genomes clustered in agreement with their phylogeny. However, strain Hudgins genome clustered away from the ZBW5 genome (Fig. 3A). A closer look at the COG function profile of S. hominis strain Hudgins in comparison to only S. hominis strains is shown in Table S1. Further analyses using genomic features (including the genome size, the number of genes, the number of transporters identified, the GC content, the number of non-coding bases, the number of genes belonging to COG categories, as well as the number of genes belonging to each COG category) cluster Staphylococcus hominis strain Hudgins genome close to S. hominis genomes SK119, RIT-PI-K, C80 and VCU122 and away from the ZBW5 genome, likely due to the enrichment in the number of transporters identified in the former genomes compared to the ZBW5 genome (Fig. 3B). Compared to other S. hominis genomes, strain Hudgins genome shared 97.8–99.4% genome average nucleotide identity and 0.89–0.93 alignment fraction. Using NsSimScan, 1920–1986 bidirectional hits were identified representing 84.6–87.5% of the total genes in the Hudgins genome (Table S2). These data are typical of strains belonging to the same species [16,17]. Blast gene homology comparisons also revealed that the majority of the genes in the Hudgins genome (84–87% of the total number of genes) were ≥95% similar to genes in other S. hominis genomes (Table S3).

4. Conclusions

This study presents the draft genome sequence and annotation of Staphylococcus hominis strain Hudgins. The genome revealed extensive sugar and amino acid degradation machinery. Comparison to the virulence factor database identified 475 genes in the genome with potential virulence-associated function including type IV Com system for pilus production, and the potential for the secretion of a membrane associated poly-gamma-glutamate capsule. Comparative genomics using general genomic features as well as the COG function profile coincided with the phylogenetic topology but placed strain Hudgins away from the previously published S. hominis genome ZBW5.

Competing interests

All authors declare no competing interests.

Funding

Staphylococcus hominis strain Hudgins was selected for sequencing as part of a project at Oklahoma State University funded by a grant from the Howard Hughes Medical Institute through the Science Education Program. The Student Initiated Microbial Discovery (SIMD) project is an undergraduate retention program that was introduced in 2015 with the aim of engaging undergraduates in authentic lab research. As part of the project, we encourage students to participate in a two-semester long research endeavor where undergraduate students isolate an environmental strain during the first semester in an introductory

---

Table 4: Genome statistics.

| Attribute                      | Value  | % of total |
|--------------------------------|--------|------------|
| Genome size (bp)               | 2,211,863 | 100.00%    |
| DNA coding (bp)                | 1,916,307 | 86.64%     |
| DNA G + C (bp)                 | 692,578  | 31.31%     |
| DNA scaffolds                   | 14      | 100.00%    |
| Total genes                    | 2270    | 100.00%    |
| Protein coding genes           | 2174    | 95.77%     |
| RNA genes                      | 96      | 4.23%      |
| Pseudo genes                   | 0       | 0          |
| Genes in internal clusters     | 417     | 18.37%     |
| Genes with function prediction | 1795    | 79.07%     |
| Genes assigned to COGs         | 1669    | 73.52%     |
| Genes with Pfam domains        | 1878    | 82.73%     |
| Genes with signal peptides     | 61      | 2.69%      |
| Genes with transmembrane helices| 556     | 24.49%     |
| CRISPR repeats                 | 1       | –          |

The table is based on the total number of protein coding genes in the genome.

Table 5: Number of genes associated with general COG functional categories.

| Code | Value | %age  | Description                                |
|------|-------|-------|-------------------------------------------|
| J    | 185   | 9.94% | Translation, ribosomal structure and biogenesis |
| A    | 0     | 0.00% | RNA processing and modification            |
| K    | 125   | 6.72% | Transcription                              |
| L    | 93    | 5.00% | Replication, recombination and repair       |
| B    | 0     | 0.00% | Chromatin structure and dynamics           |
| D    | 26    | 1.40% | Cell cycle control, cell division, chromosome partitioning |
| V    | 36    | 1.93% | Defense mechanisms                         |
| T    | 64    | 3.44% | Signal transduction mechanisms             |
| M    | 97    | 5.21% | Cell wall/membrane biogenesis              |
| N    | 4     | 0.21% | Cell motility                              |
| U    | 18    | 0.97% | Intracellular trafficking and secretion     |
| O    | 71    | 3.82% | Posttranslational modification, protein turnover, chaperones |
| C    | 100   | 5.37% | Energy production and conversion           |
| G    | 117   | 6.29% | Carbohydrate transport and metabolism      |
| E    | 169   | 9.08% | Amino acid transport and metabolism        |
| F    | 78    | 4.19% | Nucleotide transport and metabolism        |
| H    | 124   | 6.66% | Coenzyme transport and metabolism          |
| I    | 76    | 4.08% | Lipid transport and metabolism             |
| P    | 134   | 7.2%  | Inorganic ion transport and metabolism     |
| Q    | 35    | 1.88% | Secondary metabolites biosynthesis, transport and catabolism |
| R    | 165   | 8.87% | General function prediction only            |
| S    | 132   | 7.09% | Function unknown                           |
| –    | 601   | 26.48%| Not in COG                                |

The total is based on the total number of protein coding genes in the genome.
Fig. 3. (A) COG profile clustering of the genomes compared in this study. (B) Principal component analysis biplot of the genomic features and COG category distribution in the genomes compared. Genomes are represented by stars and are depicted by strain name. Arrows represent genomic features or COG categories used for comparison. The arrow directions follow the maximal abundance, and their lengths are proportional to the maximal rate of change between genomes. The first two components explained 75% of variation. Color coding: S. haemolyticus, black; S. hominis, red. Strain Hudgins is shown in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
microbiology course, followed by extraction of its genomic DNA and analysis of its genome sequence during the second semester in an upper division microbial genomics course. Here, strain Hudgins was isolated by an undergraduate student (GCH) and its genome was analyzed by a team of undergraduate (CJ and JZ) and graduate (SC) students. This is Draft Genome #8 in the SIMD project. WDH acknowledges support by NSF grants MCB-1051590, MRI-1338097, and CHE-1412500.

Authors’ contributions

SC, CJ, JZ, MBC, and NY contributed to the analysis. SC, WDH, DPF, and NY wrote the manuscript. GCH, CB, and RAH performed the lab experiments.

Transparency document

The Transparency document associated to this article can be found in the online version.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2016.10.003.

References

[1] M.B. Couger, A. Hurlbut, C.L. Murphy, C. Budd, D.P. French, W.D. Hoff, M.S. Elshahed, Y. NIH, Draft Genome Sequence of the Environmental Isolate Chryseobacterium sp. Hurlbut01. Genome Announc. 3 (5) (2015).
[2] C. Callewaert, F.-M. Kerckhof, M.S. Granitsiotis, M. Van Gele, T. Van de Wiele, N. Boon, Characterization of Staphylococcus and Corynebacterium clusters in the human axillary region. PLoS One 8 (6) (2013), e70538.
[3] R. Coates, J. Moran, M.J. Horsburgh, Staphylococci: colonizers and pathogens of human skin. Future Microbiol 9 (1) (2014) 75–91.
[4] P.M. Swe, M. Zakrzewski, A. Kelly, L. Krause, K. Fischer, Scabies mites Alter the skin microbiome and promote growth of opportunistic pathogens in a porcine model. PLoS Negl. Trop. Dis. 8 (5) (2014), e2897.
[5] K. Dansye, L. Stratton, B.D. Park, Staphylococcus Hominis carotid artery infection with septic embolization. J. Vasc. Surg. Cases 1 (2) (2015) 81–83.
[6] F. Chaves, M. Garcia-Álvarez, F. Banx, C. Alba, J.R. Otero, Nosocomial spread of a Staphylococcus Hominis subsp. NovoBiosepticus strain causing sepsis in a neonatal intensive care unit. J. Clin. Microbiol. 43 (9) (2005) 4877–4879.
[7] S. Jiang, B. Zheng, W. Ding, L. Lv, J. Ji, H. Zhang, Y. Xiao, L. Li, Whole-genome sequence of Staphylococcus hominis, an opportunistic pathogen. J. Bacteriol. 194 (17) (2012) 4761–4762.
[8] G.L. Hughes, J.A. Raygoza Garay, V. Koundal, J.L. Raffon, M. MM, Genome Sequences of Staphylococcus hominis Strains ShA31, ShA32, and ShA33. Isolated from the Asian Malaria Mosquito Anopheles stephensi. Genome Announc. 4 (2) (2016).
[9] K. Marimuthu, Isolation and characterization of Staphylococcus hominis JX061712 from oil contaminated soil. J. Pharm. Res. 7 (3) (2013) 252–256.
[10] D.R. Zerbino, E. Birney, Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18 (5) (2008) 821–829.
[11] D. Hyatt, G.L. Chen, P.F. Locascio, M.J. Land, F.W. Larimer, L.J. Hauser, Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinf. 11 (2010) 119.
[12] J. Mistry, R.D. Finn, S.R. Eddy, A. Bateman, M. Punta, Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. 41 (12) (2013), e121.
[13] R.D. Finn, P. Coggill, R.Y. Eberhardt, S.R. Eddy, J. Mistry, A.L. Mitchell, S.C. Potter, M. Punta, M. Quereshi, A. Sangrador-Vegas, et al., The Pfam protein families database: Towards a more sustainable future. Nucleic Acids Res. 44 (D1) (2016) D279–D285.
[14] N.J. Varghese, S. Mukherjee, N. Ivanova, K.T. Konstantinidis, K. Mavrommatis, N.C. Kyripides, A. Pati, Microbial species delineation using whole genome sequences. Nucleic Acids Res. (2015).
[15] V. Novichkov, A. Kaznadzey, N. Alexandrova, D. Kaznadzey, NSimScan: DNA comparison tool with increased speed, sensitivity and accuracy. Bioinformatics (2016).
[16] D. Hyatt, G.L. Chen, P.F. Locascio, M.J. Land, F.W. Larimer, L.J. Hauser, Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinf. 11 (2010) 119.
[17] J. Mistry, R.D. Finn, S.R. Eddy, A. Bateman, M. Punta, Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. 41 (12) (2013), e121.
[18] S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33 (7) (2016) 1870–1874.
[19] D. Roberts, labdev: Ordination and Multivariate Analysis for Ecology. 2007.
[20] C.R. Carlson, T.L. Denamur, V. Iakimov, C. Mailhot, R. Sirot, V. Novichkov, A. Kaznadzey, N. Alexandrova, D. Kaznadzey, Wolf PSORT: protein localization predictor. Nucleic Acids Res. 35 (Web Server issue) (2007) W585–W587.
[21] C. Rankie, P. Schweintek, A. Sczyzka, N.H. Ivanova, I.J. Anderson, F. Cheng, A. Darling, S. Malfatti, R.K. Swan, E.A. Gies, et al., Insights into the phylogeny and coding potential of microbial dark matter. Nature 499 (7459) (2013) 431–437.
[22] D. Hyatt, G.L. Chen, P.F. Locascio, M.J. Land, F.W. Larimer, L.J. Hauser, Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinf. 11 (2010) 119.
[23] D.R. Finn, P. Coggill, R.Y. Eberhardt, S.R. Eddy, J. Mistry, A.L. Mitchell, S.C. Potter, M. Punta, M. Quereshi, A. Sangrador-Vegas, et al., The Pfam protein families database: Towards a more sustainable future. Nucleic Acids Res. 44 (D1) (2016) D279–D285.
[24] N.J. Varghese, S. Mukherjee, N. Ivanova, K.T. Konstantinidis, K. Mavrommatis, N.C. Kyripides, A. Pati, Microbial species delineation using whole genome sequences. Nucleic Acids Res. (2015).
[25] V. Novichkov, A. Kaznadzey, N. Alexandrova, D. Kaznadzey, NSimScan: DNA comparison tool with increased speed, sensitivity and accuracy. Bioinformatics (2016).
[26] D. Hyatt, G.L. Chen, P.F. Locascio, M.J. Land, F.W. Larimer, L.J. Hauser, Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinf. 11 (2010) 119.
[27] J. Mistry, R.D. Finn, S.R. Eddy, A. Bateman, M. Punta, Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. 41 (12) (2013), e121.
[28] S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33 (7) (2016) 1870–1874.
[29] D. Roberts, labdev: Ordination and Multivariate Analysis for Ecology. 2007.
[30] C. Rankie, P. Schweintek, A. Sczyzka, N.H. Ivanova, I.J. Anderson, F. Cheng, A. Darling, S. Malfatti, R.K. Swan, E.A. Gies, et al., Insights into the phylogeny and coding potential of microbial dark matter. Nature 499 (7459) (2013) 431–437.
[31] D. Hyatt, G.L. Chen, P.F. Locascio, M.J. Land, F.W. Larimer, L.J. Hauser, Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinf. 11 (2010) 119.
[32] J. Mistry, R.D. Finn, S.R. Eddy, A. Bateman, M. Punta, Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. 41 (12) (2013), e121.
[33] S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33 (7) (2016) 1870–1874.
[34] D. Roberts, labdev: Ordination and Multivariate Analysis for Ecology. 2007.
[35] V. Novichkov, A. Kaznadzey, N. Alexandrova, D. Kaznadzey, NSimScan: DNA comparison tool with increased speed, sensitivity and accuracy. Bioinformatics (2016).
[36] D. Hyatt, G.L. Chen, P.F. Locascio, M.J. Land, F.W. Larimer, L.J. Hauser, Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinf. 11 (2010) 119.
[37] J. Mistry, R.D. Finn, S.R. Eddy, A. Bateman, M. Punta, Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. 41 (12) (2013), e121.