Supplementary Table S1. Methods for studying prokaryotic evolution, level of cohesion and species delineation

Purpose	Method	Advantages	Disadvantages	Influence of HGT	References
Phylogenetic cohesion	16S rRNA phylogenies	This gene is universally present and different regions of it evolves at different rates. Additionally, the popularity of using this gene provides a multitude of data across the vast diversity of prokaryotic taxa.	The gene is often too conserved, thus lacking resolving power among certain taxa. This gene has been found previously to be transferrable through HGT, and has even been found on plasmids. It is also not unusual for multiple polymorphic copies of the gene to occur within a single genome. Furthermore, a large number of 16S gene sequences in databases have been shown to be chimeric which introduces errors when identifying isolates.	An incorrect evolutionary hypothesis would be reconstructed if this gene has undergone HGT.	Boucher et al., 2004 [142]; Garrity et al., 2005 [115]; Gevers et al., 2005 [116]; Konstantinidis and Tiedje, 2007 [117]; Anda et al., 2015 [143]; Glaeser and Kämpfer, 2015 [149]
	Multi-Locus Sequence Analysis	This approach employs the analysis of multiple housekeeping genes (4 to 8) for phylogenetic analysis. The use of multiple loci overcomes problems associated with analysing a single gene (e.g., like lack in resolution and the effects of HGT). Using a limited number of genes means that sequencing efforts are cheaper and more easily attainable through routine PCR and sequencing.	The same set of MLSA genes cannot be used across all taxa. Additionally, for species delineation, it is sometimes unclear where species boundaries should be implemented and other approaches are required for better informed decisions. It is also not uncommon to observe incongruence between the individual genealogies or between phylogenies based on either the nucleotide or protein sequences.	As multiple genes are analysed, this approach is more robust in dealing with HGT than single gene phylogenies, as analysing multiple genes are thought to overcome signal associated with HGT. However, due to limited signal together with HGT in these datasets, one can often obtain an unreliable evolutionary hypothesis.	Gevers et al., 2005 [116]; Glaeser and Kämpfer, 2015 [149]; Venter et al., 2017 [114]

	Sequence similarity of marker genes	Initially, 16S rRNA gene sequence similarity was considered and was later supplemented with MLSA gene sequence similarity with cut-off values being 98.7% for 16S and ranging for MLSA genes depending on the variability of the genes employed. Information regarding marker genes is widely available and easily obtainable through routine amplification experiments.	Differences in evolutionary rate across taxa will greatly influence the inferred species boundaries, as separately, albeit slowly, evolving lineages will unavoidably be delineated as belonging to the same species, where members of a fast evolving species may be delineated as distinct species.	Using sequence similarity as species boundary cut-offs may result in taxa being considered as conspecifics due to a recent HGT event between these taxa.	Tindall et al. 2010 [133]; Peeters et al., 2013 [141]
	Extended MLSA and whole genome phylogenies	Increased character sampling associated with this approach can overcome stochastic error in the dataset, yielding more robust, well supported phylogenies.	By increasing character sampling, without representative taxon sampling, one can increase systematic error within the analysis. As these genes form part of the core genome and thought essential for survival, some may be evolving under very strong purifying selection and may lack signal for vertical inheritance. Additionally, with the increase in genes analysed there are more potential for conflicting signal within the dataset. Furthermore, the costs of sequencing an extended set of genes in multiple taxa or sequencing whole genomes is also a drawback associated with this approach.	It is thought that despite the inevitability of some genes being prone to HGT, the overall signal of vertical descent within the extended MLSA datasets will overcome this noise, thus allowing reconstruction of robust evolutionary histories.	Daubin et al., 2002 [98]; Heath et al., 2008 [145]; Philippe et al., 2011 [146]; Bennett et al., 2012 [147]; Yutin et al., 2012 [148]; Palmer et al., 2017 [126]

	Phylogenetic networks	This approach does not force a bifurcating topology on the data. It also depicts all signals present in the dataset; i.e., signal for vertical inheritance together with all conflicting signal (including those associated with HGT).	These analyses can be computationally intensive and various network approaches cannot yet easily be implemented for genome scale data.	As all signals present within the dataset is depicted, both vertical and horizontal evolutionary processes can be accommodated.	Bryant and Moulton, 2004 [122]; Huson and Bryant, 2006 [123]
	Genealogical concordance	The consensus signal obtained from multiple individual gene trees provides a means to infer species boundaries. The species boundary is considered to be the point between reticulation and divergence. Often, sequence data for multiple gene loci are available in the form of MLSA data.	Multiple gene sequences are required to obtain individual genealogies.	As individual genealogies are considered, genes with conflicting topologies can be identified. It is however, difficult to distinguish between incomplete lineage sorting and HGT for these conflicting topologies. The majority consensus of the genealogies is then considered as the basis of taxonomic decisions.	Avise, 1990 [129]; Taylor et al., 2000 [130]; Dettman et al., 2003 [131]; Sarver et al., 2011 [132]; Venter et al., 2017 [114]
Phenotypic cohesion	Physiological differences	To investigate cohesion in terms of phenotypes as the basis for species delineation, taxa are often considered in terms of similarities in metabolic and physiological characteristics. These tests comprise mostly of fatty acid analyses and various biochemical tests, like the utilization of various carbon	The increased sampling of diversity within various genomically cohesive groups has indicated the immense variability observed within groups for these tests and these tests are no longer informative. The results of this approach were for differentiation purposes and not for phenotypic characterization. Additionally, expression differences during *in vitro* testing can result in false negatives.	Some of these traits can be mediated through a single gene, thus these traits can be horizontally acquired. This means that the presence of specific characteristics does not necessarily mean that individuals inherited these traits vertically.	Stackebrandt et al., 2002 [134]; Sutcliffe et al., 2012 [144]; Sutcliffe, 2015 [140]

		sources and amino acids. This approach provides initial indications of metabolic capacity and can give an indication of relatedness. There are various systems that have been developed to aid in screening for traits.	Many of these characteristics are also thought to be situated on the accessory genome and may not be informative for taxonomic purposes in a system based on the shared core genome.		
	Functional annotation	This method allows functional annotation of genes through inferred orthology to experimentally verified functions of genes. A large number of genes can be annotated simultaneously and various platforms for accessing these databases are available. Known metabolic networks can also be reconstructed to obtain a clearer picture of the metabolic potential of taxa and allow comparisons of predicted biology, which in turn allow inferences regarding phenotypic cohesion.	Whole genome sequence data is required to apply this methodology. As only functional inferences to experimentally verified genes can be done, a large proportion of the genome goes unannotated. This is due to many genes not being functionally characterized and remains denoted as hypothetical or unknown genes. Additionally, the absence of specific orthologous genes does not necessarily mean the absence of a specific metabolic pathway or process, as alternate, as-of-yet uncharacterised pathways may exist and can potentially be employed by the organism to perform the same function.	As functions for genes are inferred, horizontally acquired traits within individuals of the same species can be identified. Also, the gains or losses of specific traits in members of a species can be inferred for further investigations.	Kanehisa, 2000 [121]; Kanehisa *et al.*, 2016 [120]; Palmer et al., 2018a [127], b [128]
Genomic cohesion	DNA-DNA Hybridisation	The basic principle of this approach is based on the assumption that highly similar genomes will hybridize with one another	There is limited repeatability, particularly between laboratories. This approach is also very labour-intensive and extensive experience is required to produce reliable results. A	DDH values would be lower between taxa that have undergone extensive HGT. However, chimeric individuals may hybridise	Stackebrandt et al., 2002 [134]; Goris et al., 2007 [135]; Achtman and Wagner, 2008 [80];

		more efficiently than to less similar genomes. A DDH value ≥70% and a melting temperature difference of less than 5°C is thought to be characteristic of conspecifics.	large amount of high-quality genomic DNA is required for each experiment. Differences in genome size can dramatically affect reciprocal results. Cumulative databases for results for future comparisons cannot be constructed. The appropriateness of cut-off values is questionable, as the value was determined from re-association experiments from predetermined species.	very efficiently to the reference at some genomic regions (at sufficient levels to be conspecifics), but not at other regions. For example, individuals of the same species may have low overall DDH values due to the presence of large genomic islands.	Richter and Rosselló-Móra, 2009 [136]; Meier-Kolthoff et al., 2013 [137]; Steenkamp et al., 2015 [139]
	Overall Genome Relatedness Index	The sequence similarity of the entire genome can be brought into consideration. The most widely applied approach is Average Nucleotide Identities (ANI) and *in silico* DDH. These methods provide an indication of the overall genomic cohesion between taxa. Software for calculation of these metrics is freely available and user-friendly.	Lineage-specific evolutionary rate heterogeneity will influence the reliability of these metrics to denote species boundaries. If large fragments of DNA below the homology thresholds are present within a taxon, these fragments would be excluded from the analysis and an over estimate of the genomic similarity will be obtained. The proportion of the genome considered to be homologous thus need to be reflected in these calculations, thus further development of this approach is in progress. Also, though it has been suggested that relatively robust calculations can be made using partial genomes, whole genome sequences are generally required.	Horizontally acquired fragments below the homology thresholds would be discarded from the analysis. If these fragments are above the thresholds, the effect of HGT would be averaged out across the genomic regions analysed. Specific information regarding HGT is thus not attainable.	Goris et al., 2007 [135]; Konstantinidis and Tiedje, 2007 [117]; Richter and Rosselló-Móra, 2009 [136]; Varghese et al., 2015 [138]