**Prompt:** *Discuss the challenges involved in defining a microbial species and how HGT complicates matters, especially in the context of the evolution and phylogenetic distribution of microbial metabolic pathways. Can you comment on how HGT influences the maintenance of global biogeochemical cycles through time? Finally, do you think it is necessary to have a clear definition of a microbial species? Why or why not?*

Defining a species at the microbial level is difficult given their relatively small genome size, which can range from a mere 0.2 to a significantly larger 13.0 megabases, depending on the microbe’s lifestyle (1, 2). Using *E. coli* as an example, the sequence similarity required to be classified the same species has been shown to vary greatly, complicating the definition of a species (3) Horizontal Gene Transfer can also complicate these phylogenetic relationships between organisms, creating incongruent relationships, whereby there are no partitions/taxonomic groups in common (4). As microbial genome sizes are small, they are under evolutionary pressure to become more streamlined, losing genes in the process (5) However, these genes that are selectively lost from the organism are not lost from the community, as they can be transferred to other organisms via Mobile Genetic Elements (MGE) (6). Gene composition changes in microbial genomes through Horizontal Gene Transfer can accumulate over time with changing environmental evolutionary pressures over time, which can change the biogeochemical landscape of ecosystems as specific pathway genes are gained and lost (7.).

Horizontal Gene Transfer is the driver of microbial evolution, where it can alter the relatively small microbial genome of a certain species, leading to microbial strains of different lengths and consisting of drastically different genes (3, 4). As these genomes can potentially be quite distinct from each other in the pathways that are expressed, there is the challenge of having a clear definition of a microbial species where one strain can be very different functionally. For example, in Welch *et al.*, 2002, the sequence similarity between the three *E. coli* strains sequenced in the study was found to be 39.2%, which is significantly lower than the 99.5% sequence similarity between human genomes (8). In the context of *E. coli*, each strain cancontain different pathogenicity islands in their genomes that can alter their disease potential, and also their host range based on selective pressures in the environment (3). In addition, antibiotic resistance provides another avenue of selective pressure on *E. coli* strains, where only the resistant bacteria will survive in the presence of a certain antibiotic that they are resistant to. These antibiotic resistance genes primarily come from Horizontal Gene Transfer, through which it can be rapidly acquired as it would be required for survival when the antibiotic is present (9).

However, in the context of biogeochemical cycling, the selective pressures lean towards the energetic and nutritional requirements of microbes that can change under stress from environmental perturbations, where genes are transferred to other microbes in the community based on which pathways are favourable in a given environment (9). These pathways can be distributed across a wide range of microbes as they were drawn from a primitive global gene pool (9). The influence imposed on genes by the environment also favours certain genes that are present across different microbial species, where the favourable genes are transferred throughout the community, such as the transfer of methanogenesis, sulfur reduction, and (de)nitrification genes (9). Furthermore, if the certain pathway continues to be functionally and energetically favourable in the environment due to a lack of biogeochemical changes in the environment, then it’s retention would be also be governed by environmental selective pressures that could affect the survival of members within a microbial community. This could differentially distribute biogeochemical pathways between communities, such as in the oceans where nutrient and oxygen concentrations vary with depth (10, 11). Similarly, different microhabitats exist within hydrothermal vent communities, where the distance from the vent can drastically change the biogeochemical pathways as the environment becomes less harsh and reducing (11, 12).

Horizontal Gene Transfer has been suspected due to changes in codon usage frequencies and drastically different genes were found in different pathogenicity islands that are inserted at certain regions in the genome (3). As genes of different lengths are inserted, this can affect genome length differently, depending on whether it is a single gene, or an entire pathway that is inserted or substituted, which can reduce the similarity between the different strains of *E. coli*. Furthermore, there could be function-dependent biases for certain genes that increase the rate of Horizontal Gene Transfer, as those genes could possibly confer an advantage, whether it be in the ecology or pathogenicity of a microbe (13). Analysis of Epsilon-proteobacterial genomes has shown that hydrothermal vent microbes can have very similar virulence genes as their human/animal pathogenic counterparts, but their ability to easily acquire new genes causes them to diverge based on selective pressures in the environment they inhabit (14).

Genes that are acquired by Horizontal Gene Transfer can be accumulated over time through insertions and substitutions that could increase the activity of a certain pathway or introduce new genes and/or pathways. This could possibly confer a fitness advantage to the microbe as it creates a niche, or increase its abundance within the community. However, this should come at the cost of other unused genes given that organisms that accept genes via Horizontal Gene Transfer generally have a smaller genome, which is the result of genes being lost (5). Furthermore, selecting for certain genes allows the genome to be more streamlined, especially useful in the ocean, where resources can be limiting in certain regions (5). Over time, when the new horizontally acquired genes accumulate, new pathways emerge in a community, and only the pathways that can survive the environmental perturbations that appear over time will be conserved and be distributed to other organisms in the community. These environmental perturbations can cause changes in larger scale biogeochemical cycles that will add up and cause genes to diverge over longer time scales (5).

The mechanisms that can be used by Horizontal Gene Transfer include transformation, conjugation and transduction, in which Mobile Genetic Elements (MGEs) help facilitate. MGEs can include plasmids, bacteriophages, transposons and integrative conjugative elements (ICEs) that can move DNA from one spot to another on the genome or to other organisms that not be closely related (6,15). Prokaryotes use an operon to organize the genes with a single promoter at the start, which are highly conserved. However, increasing the distance from the promoter increases the likelihood of gene insertion (6). This creates hotspots where new genes are more likely to be integrated to the genome whether it be through the introduction of prophages or the ICEs that can be several hundred base pairs long, and can be self managed using the genes encoded within them (6).

As genome lengths can vary drastically within the same species, this complicates the process of defining the criteria for a microbial species, even with a conserved set of genes within a species. It is thought that genome lengths can affect the number of MGEs and ICEs, increasing Horizontal Gene Transfer rates with a larger genome size (2). Furthermore, as larger genomes have more genes that originated from Horizontal Gene Transfer, this would also increase the degree of phylogenetic incongruence as there are now more foreign genes incorporated (2). Also, a larger, more complex genome can also be more modular, increasing organization and reducing the likelihood of deletions, leading to more successful Horizontal Gene Transfers and integration (2).

Phylogenetic incongruence from Horizontal Gene Transfer can also introduce a source of complexity when trying to determine species relationships on the phylogenetic tree, where a gene can jump from one branch/leaf to another, distorting the phylogenetic relationships on the tree. This can make it hard identifying species similarity, as similar species are now in a different branch on the tree where they share the same sequences (4). Although microbes are connected to each other on the branch of the phylogenetic tree, even if they have sequence similarity by virtue of being the same species, doesn’t not guarantee they would have the same metabolic pathways (4).

Given that Horizontal Gene Transfer can drastically alter the genetic makeup within different strains of a species, it may seem difficult and impossible to determine a single definition to differentiate species, especially in the context of microbial genomes. However, if a set of criteria defining what constitutes a species can be used to reliably determine the conserved genes across all strains within a species, then that could be used to clearly define and differentiate microbial species. As long as two organisms share a certain degree of sequence homology that is common across all strains, then they can be classified as the same species, even being as low as 39.2% for *E. coli* (3). This would help with the classification of microbes into their species as it adds an extra filter, even Horizontal Gene Transfer increases the complexity of the microbial genome.

In conclusion, Horizontal Gene Transfer is one of the main drivers behind evolution, constantly changing microbial genomes through evolutionary selective pressures that can introduce new pathways or conserve existing ones. This also has the power to overhaul the genome sizes of certain strains by way of the insertion of MGEs and/or ICEs into the genome at higher frequencies certain hotspots, as well as the deletion of unused genes (6). This all decreases the sequence similarity thresholds for microbial species as native sequences are replaced by foreign ones inserted by Horizontal Gene Transfer and increases the within-species variability. With a higher genome complexity and within-species variability also comes with higher uncertainty with phylogenetic relations as phylogenetic incongruence increases. Therefore, an universally accepted standard for differentiating microbes at the species level should be the aim, where an exact percentage threshold is not used. This would make it easier to form phylogenetic relations between strains and have a solid definition to base future research on. Also, further research can be done on how genome size, in conjunction with gene function, affects the rate of Horizontal Gene Transfer within a community (2), so we can better understand how they are selected based on their biogeochemical activity in the environment.

**References:**

1. Koonin EV, Wolf YI. 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res. 36(21): 6688-6719.
2. Cordero OX, Hogeweg P. 2009. The impact of long-distance horizontal gene transfer on prokaryotic genome size. Proc Natl Acad Sci. 106(51): 21748-21753.
3. Welch RA, Burland V, Plunkett III G, Redford P, Roesch P, Rasko D, Buckles EL, Liou SR, Boutin A, Hackett J, D Stroud, Mayhew GF, Rose DJ, Zhou S, Schwartz, DC, Perna NT, Mobley HLT, Donnenberg MS, Blattner FR. 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc Nat Acad Sci. 99(26): 17020-17024.
4. Philippe H, Douady CJ. 2003. Horizontal gene transfer and phylogenetics. Curr Opin Microbiol. 6(5): 498-505.
5. Morris JJ, Lenski RE, Zinser ER. 2012. The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss. MBio. 3(2): e00036-12. doi: 10.1128/mBio.00036-12
6. Oliveira PH, Touchon M, Cury J, Rocha EPC. 2017. The chromosomal organization of horizontal gene transfer in bacteria. Nat Commun. 8(841): 1-11.
7. McDaniel LD, Young E, Delaney J, Ruhnau F, Ritchie KB, Paul JH. 2010. High Frequency of Horizontal Gene Transfer in the Oceans. Science. 330(6000): 50.
8. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AWC, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, Strausberg RL, Venter JC. 2007. The Diploid Genome Sequence of an Individual Human. PLoS Biol. 5(10): 2113-2144.
9. Falkowski PG, Fenchel T, Delong EF. 2008. The Microbial Engines That Drive Earth's Biogeochemical Cycles. Science. 320(5879):1034-1039. doi:[10.1126/science.1153213](https://doi.org/10.1126/science.1153213)
10. Xie W, Wang F, Guo L, Chen Z, Sievert SM, Meng J, Huang G, Li Y, Yan Q, Wu S, Wang X, Chen S, He G, Xiao X, Xu A. 2011. Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. ISME J. 5(3): 414-426.
11. Cerqueira T, Barroso C, Froufe H, Egas C, Bettencourt R. 2018. Metagenomic Signatures of Microbial Communities in Deep-Sea Hydrothermal Sediments of Azores Vent Fields.
12. Vetriani C, Voordeckers JW, Crespo-Medina M, O’Brien CE, Giovannelli D, Lutz RA. 2014. Deep-sea hydrothermal vent Epsilonproteobacteria encode a conserved and widespread nitrate reduction pathway (Nap). ISME J. 8(7): 1510. doi: 10.1038/ismej.2013.246
13. Nakamura Y, Itoh T, Matsuda H, Gojobori T. 2004. Biased biological functions of horizontally transferred genes in prokaryotic genomes. Nat Genet. 36(7): 760-766.
14. Nakagawa S, Takaki Y, Shimamura S, Reysenbach AL, Takai K, Horikoshi K. 2007. Deep-sea vent ε-proteobacterial genomes provide insights into emergence of pathogens. Proc Natl Acad Sci. 104(29): 12146-12150
15. Koonin EV, Makarova KS, Aravind L, Frost LS, Leplae R, Summers AO, Toussaint A. 2005. Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol. 3(9): 722-732.