Prokaryote pangenomes are dynamic entities

Cummins, Elizabeth A; Hall, Rebecca J; McInerney, James O; McNally, Alan

DOI:
10.1016/j.mib.2022.01.005

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Cummins, EA, Hall, RJ, McInerney, JO & McNally, A 2022, 'Prokaryote pangenomes are dynamic entities', *Current Opinion in Microbiology*, vol. 66, pp. 73-78. https://doi.org/10.1016/j.mib.2022.01.005

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

Prokaryote pangenomes are dynamic entities

Elizabeth A Cummins[1,3], Rebecca J Hall[1,3], James O McInerney[2] and Alan McNally[1]

Prokaryote pangenomes are influenced heavily by environmental factors and the opportunity for gene gain and loss events. As the field of pangenome analysis has expanded, so has the need to fully understand the complexity of how eco-evolutionary dynamics shape pangenomes. Here, we describe current models of pangenome evolution and discuss their suitability and accuracy. We suggest that pangenomes are dynamic entities under constant flux, highlighting the influence of two-way interactions between pangenome and environment. New classifications of core and accessory genes are also considered, underscoring the need for continuous evaluation of nomenclature in a fast-moving field. We conclude that future models of pangenome evolution should incorporate eco-evolutionary dynamics to fully encompass their dynamic, changeable nature.

Addresses
[1] Institute of Microbiology and Infection, College of Medical and Dental Sciences, University of Birmingham, Birmingham, B15 2TT, UK

[2] School of Life Sciences, University of Nottingham, Nottingham, NG7 2UH, UK

Corresponding authors: Hall, Rebecca J (r.j.hall@bham.ac.uk), McNally, Alan (a.mcnally.1@bham.ac.uk)
[3] These authors contributed equally and are listed alphabetically by surname.

Current Opinion in Microbiology 2022, **66**:73–78

This review comes from a themed issue on **Environmental microbiology**

Edited by **Paul Hoskisson** and **John Bruce**

For complete overview of the section, please refer to the article collection, "Environmental Microbiology"

Available online 31st January 2022

https://doi.org/10.1016/j.mib.2022.01.005

Introduction

The field of pangenomics emerged with the acceleration of genome sequencing and the observation of intra-specific genomic variation. It is therefore a young, rapidly expanding area of research. The multi-genome perspective required for pangenomics means that pangenome evolution is heavily nuanced with the complexity of prokaryote genome evolution and is itself a nascent concept. Genome evolution is driven by gene acquisition via horizontal gene transfer (HGT), as well as extensive gene loss [1,2], both of which subsequently shape pangenomes. Multiple models of pangenome origin and maintenance have been proposed [3–5] that consider different evolutionary mechanisms and account for different variables such as random drift, effective population size, selection coefficients, and HGT. Contributions of adaptive and neutral evolution in pangenomes, based on neutral and nearly neutral population genetic theory, and how they can account for extensive accessory genomes have been discussed in detail [4–10]. Some research suggests that accessory genomes are largely under the influence of neutral evolution [8]. The presence of accessory genes at the tips of trees, reflecting their transience, would be consistent with their expected distribution under the neutral model. Alternatively, others suggest that pangenomes are instead the result of adaptive evolution due to the acquisition of beneficial genes [5], with evidence suggesting dependent relationships between accessory genes [11]. The manifestation of the latter theory has been observed in a *Pseudomonas* pangenome whereby genes sharing common function were found to co-occur within the same genome more often than expected by chance, indicating adaptive structure in the accessory genome [11]. This theory of adaptive evolution in pangenomes is supported by evidence that neutral evolution did not shape the genomic diversity of *Listeria*, where no correlation was found between the average nucleotide diversity of core genes and phylogroup prevalence [12••]. A correlation may have been expected with neutral evolution under the assumption that mutations are more likely to accumulate and be maintained over time in large, multi-niche populations [13].

Incorporating ecology, including niche migration, is the next stage required to further progress in modelling pangenome evolution. Environmental preferences are more influential on pangenome evolution than inherited gene content [14•]. Modelling has already shown that accessory genes are mostly beneficial [15], but gene-by-environment interactions now need to be considered, given that fitness effects can be context-dependent [16]. The assumption of constant selection coefficients neglects the influence that niche adaption [10], frequency dependent selection [17], and fluctuating environments [18] have on pangenome evolution [19•]. These interactions between pangenome and environment are two-way. Pangenomes influence their immediate environment via the encoding of, for example, secreted public goods

including, but not limited to, metabolic by-products, siderophores, hydrolytic enzymes or even DNA [20–22]. The external environment is consequently under constant flux, and this thereby provides a different set of conditions under which the pangenome evolves. The resulting gene gain and loss, or allelic variation, may be particularly rapid when adapting to new environments [12••]. Considering pangenomes as dynamic entities may therefore better encompass the ecological nature of a species.

Pangenomes as dynamic entities

Pangenomes are considered traditionally in reasonably static terms; having 'open' or 'closed' structures depending on the number of core and accessory genes [23]. Total gene number within a pangenome is dependent on, and inextricably linked to, the number of genomes sampled. Heap's Law (Eq. (1)) was adapted from information retrieval theory [24] to model the relationship between the number of distinct genes found and the number of genomes sampled, where N is the number of gene families within the pangenome, n is the number of genomes sampled, and k and α are constants of proportionality (for fitting purposes). Simply, this formula informs that the rate at which novel genes are found decreases as genome sampling increases.

$$N = kn^{-\alpha} \quad (1)$$

The parameter α from Heap's Law is also used to define a pangenome as open or closed ($\alpha \leq 1$: closed, $\alpha > 1$: open). Open pangenomes are essentially unbounded in size whereas closed pangenomes tend toward a finite number of genes as more genomes are sampled. Open pangenomes usually coincide with greater metabolic flexibility and the occupation of multiple niches, whilst closed pangenomes indicate more ecologically specialised lifestyles [25,26••]. Every independent genome sampled in an open pangenome will provide additional genes, meaning that for many species the true pangenome may never be uncovered. Additionally, dependence on genome sampling means that all pangenome analysis performed to date may only provide a 'keyhole perspective' of the true genetic diversity of the studied organism.

Whilst useful, considering pangenomes and their subsequent evolution in relation to Heap's Law risks excluding the dynamic interplay between genome and ecology; the likelihood of gene gain and loss, the need for flexibility in certain environments over others, and the resulting impact on pangenome evolution.

Gene classification in pangenomes

As the union of all genes occurring within a set of genomes, the pangenome is often considered in terms of two major 'sections'. The inter-sectional area is referred to as the core genome and all genes outside of this intersection are classified into the accessory genome, although in practise a defined core genome is often not this strict. Shell and cloud are common subdivisions within the accessory genome that classify genes at more specific presence thresholds [27]. Shell genes are moderately conserved and present at intermediate frequencies, whereas cloud genes have very low frequencies or are uniquely present in a single genome [27]. Horesh *et al.* recently expanded upon accessory subdivisions by introducing a method to break down the accessory genome into twelve classes based on observed gene frequencies in individual lineages, rather than across the entire dataset [26••]. This non-binary approach is said to account for bias brought about by uneven sampling and disregard of population structure. The extensive subcategorisation of a pangenome in this way can highlight differing evolutionary dynamics present within an *Escherichia coli* pangenome. For instance, the ST10 and ST23 lineages had above average mean numbers of lineage-specific rare genes, shared more intermediate and rare genes between them, and possessed a high number of genes relating to genetic mobility [26••]. These findings imply that an increased ability to acquire mobile genes may drive higher levels of gene sharing between lineages, adding evidence to the notion that ST10 and ST23 are generalists that can act as genetic reservoirs for other *E. coli* lineages [28–30].

Finer division within the accessory genome also allows genes that are core to a single lineage to be considered. These genes are interesting from an ecological perspective as some phylogroup-specific core genes may provide an evolutionary advantage in a specific environment or genetic landscape. For example, the aforementioned study of *Listeria* found that isolates with a wider habitat range had more open pangenomes and a larger fraction of genes under positive selection, particularly phylogroup-specific core genes [12••]. The Horesh classification does however fail to account for 'singleton' genes uniquely present in a single genome. Singletons are interesting biologically because they may be indicative of inter-species HGT, which could for example have significance with regards to the transfer of multidrug resistance (MDR) plasmids between species or genera.

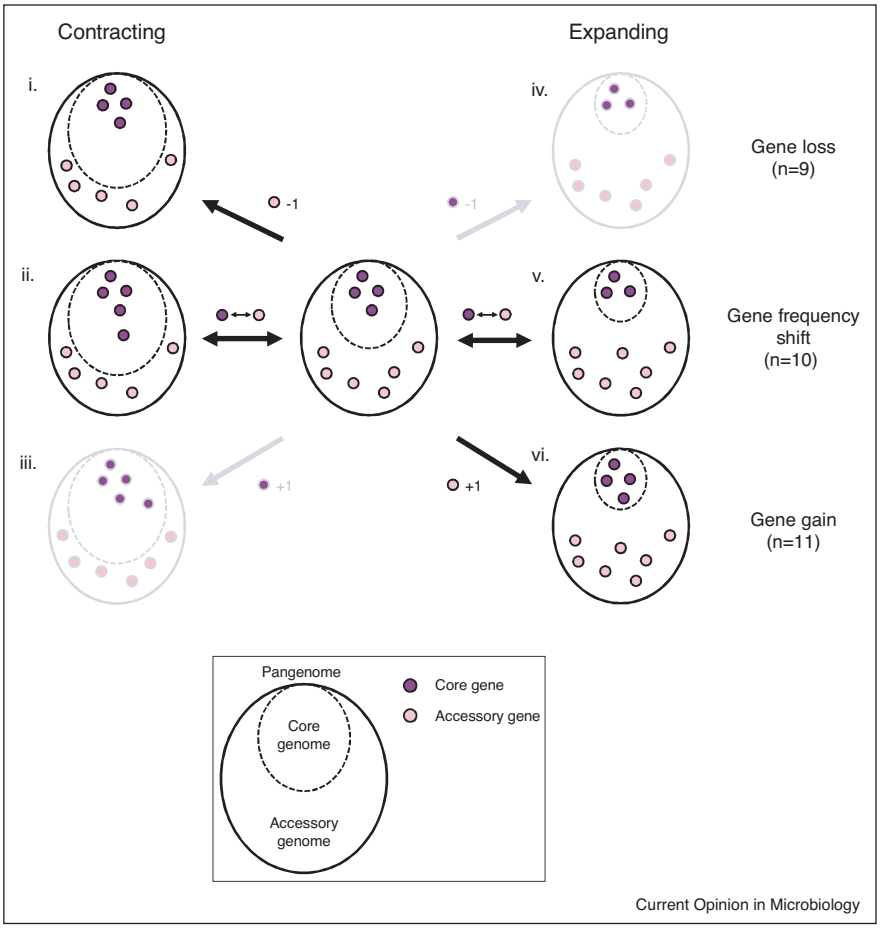
The ratio of accessory to core genome within a pangenome can be inferred by its openness. The variable nature of prokaryotic genomes, both within and between genomes, means that a gene's classification within a pangenome is not necessarily fixed. A gene once considered core to a species can become accessory [31] due to a shift in gene frequency within the population. The opposite scenario is also possible. For instance, retention of an MDR plasmid by an individual within a nosocomial outbreak could escalate to transmission throughout the population, thus encapsulating a shift in gene frequency from accessory to core. As a consequence of these

potential gene frequency shifts, the ratio of core to accessory genome is also not fixed. A caveat of current literature is that pangenomes are not typically thought of as dynamic entities, and the fluidity of gene classification is often not considered. A shift in pangenome composition (core:accessory) provides two evolutionary trajectories — to 'expand' and become more open, or 'contract' and become more closed (Figure 1).

The variable impact of gene gain and loss

Expanding and contracting are dynamic terms, used previously in the context of genome evolution [2]. They allow a consideration of pangenomes as reflective of their ecology and the relative influence of gene gain and loss events. Species in variable environments, or with a broad habitat range, may have expanding pangenomes as a result of increased opportunities for DNA acquisition, the lineage-specific loss of core genes resulting in an increasing accessory genome, or selection on pangenomes for traits that enhance flexibility. In contrast, pangenomes could be considered contracting when accessory gene loss outweighs accessory gene gain. This can occur either by the complete loss of a gene or by its fixation in the population, and thus becoming a core gene (Figure 1). Gene gain or loss may not therefore lead to an increase or decrease, respectively, in pangenome size, but rather may affect pangenome structure by altering the frequency distribution of genes within it. HGT is known to be highly influential across different genera; approximately 5.8% of the *Burkholderia pseudomallei* genome is estimated to consist of genomic islands, known hotspots of recombination [32], for example, and *E. coli* has an accessory genome influenced by mobile genetic elements [33,34]. Species within environments rich in sources of available DNA, including the marine microbe *Pseudoalteromonas*,

Figure 1



The impact of gene gain and loss within a pangenome is dependent on gene frequency distribution.
Pangenome size remains constant when events move genes between the core and accessory genome, resulting in pangenomes that become more open or closed without change in absolute size (ii, v). In other instances, gain of a novel accessory gene increases pangenome size (vi), whilst the loss of an accessory gene to the species reduces it (i). Gain (iii) or loss (iv) of core genes that alter pangenome size are considered unlikely events.

also often have pangenomes that are influenced heavily by HGT [31]. For these species, incoming DNA may expand the absolute size of the pangenome if the genes are new to the species, have no effect on the structure if the genes already exist within the accessory genome, or impact the structure if the acquisition event moves the gene from the accessory to the core genome (Figure 1). A single acquisition event may lead to a proportionally large increase in pangenome size when the incoming genes are unique to the pangenome [35]. Here, individual lineages can exploit environmental perturbations to expand within the niche and out-compete less flexible populations, and may sweep to fixation. Expanding pangenomes can also be influenced by the spatial separation of lineages. Within the *Sulfurovum* genus, genes involved in inorganic ion transport and metabolism, including four related to phosphate uptake and regulation, have been found to be significantly enriched in individuals from a lower phosphate environment [36•]. These lineage-specific genes may have been uniquely gained by one population, or, alternatively, lost from the core genome. The extent to which a pangenome can expand or contract also hinges on intrinsic factors, including barriers to HGT created by genetic elements such as the CRISPR/Cas systems encoded by *Staphylococcus lugdunensis* [37] or *Pseudomonas aeruginosa* [38]. These barriers limit not only absolute pangenome size, but also the relative movement of genes between core and accessory genome.

Perhaps counter-intuitively, expanding pangenomes also can be driven by gene loss in the case of a lineage-specific loss of a previously core gene, thereby expanding the pool of accessory genes. The *Pseudoalteromonas* pangenome exhibits species-specific gene loss events, including those involved in the metabolism of glycogen (*glyA*, *glyX*) [31]. These genes may once have been core genes, but at some point during the evolution of the genus have become accessory after loss by a subset. Differential metabolic gene presence is also influential in the *Yersina* genus. *Yersinia pestis* strain Angola retains the methionine salvage pathway *mtnKADCEBU* that is present in serovars of *Yersinia pseudotuberculosis* and *Yersinia enterocolitica* but is absent in modern *Y. pestis*, suggesting it has been lost in the latter [39].

In more restricted or constant environments with fewer gene acquisition opportunities, gene loss may influence absolute pangenome size. Host-restricted bacteria may have very little intra-specific genomic variation due to small population sizes and little opportunity for HGT [40]. This is epitomised in the 0.64 Mb genome of the obligate symbiont *Buchnera aphidicola*, within which it is predicted that no HGT has occurred for 50–70 million years whilst gene inactivation and loss has been ongoing [41]. This is an extreme example of a contracting pangenome, whereby gene loss clearly outstrips gene gain as a direct consequence of the microenvironment. These obligate symbionts contrast starkly with non-host associated organisms that generally have larger genomes [42,43], and for which the microenvironment does not select for large-scale gene loss.

Beyond the pangenome

A point of consideration is the stage at which species with very small accessory genomes cease to have definable pangenomes. A highly specialised symbiont with a nearly minimal genome may for all intents and purposes be considered as an organelle. If obligate symbionts with extremely small accessory genomes could be considered to have a pangenome, this rationale might also be extended to organelles. At this level there might be no pangenome, or rather the pangenome is equal to the core genome, but there still remains inter-genomic differences by way of allelic variation. Research in the smallest possible core genome suggests that it could comprise the approximately 53 ribosomal genes that are essential for bacterial replication [44], although here the question of whether this constitutes the true smallest pangenome is rooted perhaps in nomenclature rather than biology.

Concluding remarks

Whilst pangenomes are evolving biologically via ecological interactions and artificially with each genome sequenced, the term 'pangenome' itself is evolving to serve as a more informative definition. Dynamic definitions incorporate the knowledge that pangenomes are only snapshots in time; expanding pangenomes may contract with a small ecological change or important gain/loss event, and an 'open versus closed' label may not encompass this change fully. A gene may be present in a historical isolate, for example, but not appear in any modern genome. We suggest that pangenomes should not exist as a historical record of a species, otherwise the concept of gene loss becomes largely irrelevant. A dynamic definition allows an acknowledgement of the complement of genes that are encoded by a species at any given point in evolutionary time, whilst also considering the importance of what is biologically possible for the species today. The influence of the community on the pangenome of an individual should also be considered [45], particularly with regards to the pool of possible unique genes that could be acquired. Considering pangenomes more holistically, as dynamic entities, allows eco-evolutionary dynamics to be incorporated more completely.

Funding

EAC is funded by the Wellcome Antimicrobial and Antimicrobial Resistance (AAMR) DTP (108876B15Z). RJH is supported by a NERC grant (NE/T01301X/1), awarded to AM.

Conflict of interest statement

Nothing declared.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Treangen TJ, Rocha EPC: **Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes**. *PLoS Genet* 2011, **7**:e1001284.
2. Puigbò P, Lobkovsky AE, Kristensen DM, Wolf YI, Koonin EV: **Genomes in turmoil: quantification of genome dynamics in prokaryote supergenomes**. *BMC Biol* 2014, **12**:66.
3. Baumdicker F, Hess WR, Pfallelhuber P: **The infinitely many genes model for the distributed genome of bacteria**. *Genome Biol Evol* 2012, **4**:443-456.
4. Bobay LM, Ochman H: **Factors driving effective population size and pangenome evolution in bacteria**. *BMC Evol Biol* 2018, **10**:153.
5. McInerney JO, McNally A, O'Connell MJ: **Why prokaryotes have pangenomes**. *Nat Microbiol* 2017, **3**:1-5.
6. Shapiro BJ: *The Population Genetics of Pangenomes*. Nature Publishing Group; 2017.
7. McInerney JO, McNally A, O'Connell MJ: **Reply to 'The population genetics of pangenomes'**. *Nat Microbiol* 2017, **11**:1575.
8. Andreani NA, Hesse E, Vos M: **Prokaryote genome fluidity is dependent on effective population size**. *ISME J* 2017, **11**:1719-1721.
9. Vos M, Eyre-Walker A: **Are pangenomes adaptive or not?** *Nat Microbiol* 2017, **2**:1576.
10. Sheppard SK, Guttman DS, Fitzgerald JR: **Population genomics of bacterial host adaptation**. *Nat Rev Genet* 2018, **19**:549-565.
11. Whelan FJ, Hall RJ, McInerney JO: **Evidence for selection in the abundant accessory gene content of a prokaryote pangenome**. *Mol Biol Evol* 2021, **5**:msab139.
12. •• Liao J, Guo X, Weller DL, Pollak S, Buckley DH, Wiedmann M *et al.*: **Nationwide genomic atlas of soil-dwelling *Listeria* reveals effects of selection and population ecology on pangenome evolution**. *Nat Microbiol* 2021, **6**:1021-1030

Constructed a *Listeria* pangenome across different environments and uncovered the importance of lineage-specific core genes as drivers of adaptation.

13. Calland JK, Pascoe B, Bayliss SC, Mourkas E, Berthenet E, Thorpe HA *et al.*: **Quantifying bacterial evolution in the wild: a birthday problem for *Campylobacter* lineages**. *PLoS Genet* 2021, **17**:e1009829.
14. • Maistrenko OM, Mende DR, Luetge M, Hildebrand F, Schmidt TSB, Li SS *et al.*: **Disentangling the impact of environmental and phylogenetic constraints on prokaryotic within-species diversity**. *ISME J* 2020, **2**:1247-1259

Relative influences of phylogenetic inertia and the environment on pangenome evolution.

15. Sela I, Wolf YI, Koonin EV: **Theory of prokaryotic genome evolution**. *Proc Natl Acad Sci U S A* 2016, **113**:11399-11407.
16. Domingo-Sananes MR, McInerney JO: **Mechanisms that shape microbial pangenomes**. *Trends Microbiol* 2021, **1**.
17. Levin BR: **Frequency-dependent selection in bacterial populations**. *Philos Trans R Soc B* 1988, **7**:459-472.
18. Richardson EJ, Bacigalupe R, Harrison EM, Weinert LA, Lycett S, Vrieling M *et al.*: **Gene exchange drives the ecological success of a multi-host bacterial pathogen**. *Nat Ecol Evol* 2018, **2**:1468-1478.
19. • Sela I, Wolf YI, Koonin EV: **Assessment of assumptions underlying models of prokaryotic pangenome evolution**. *BMC Biol* 2021, **19**:27

Critical evaluation of assumptions that underpin prokaryote genome evolution in the context of pangenome dynamics.

20. McInerney JO, Pisani D, Bapteste E, O'Connell MJ: **The public goods hypothesis for the evolution of life on earth**. *Biol Direct* 2011, **6**:41.
21. Brockhurst MA, Buckling A, Racey DA, Gardner A: **Resource supply and the evolution of public-goods cooperation in bacteria**. *BMC Biol* 2008, **6**:20.
22. Smith P, Schuster M: **Public goods and cheating in microbes**. *Curr Biol* 2019, **29**:R442-R447.
23. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL *et al.*: **Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome"**. *Proc Natl Acad Sci U S A* 2005, **102**:13950-13955.
24. Heaps HS: *Information Retrieval: Computational and Theoretical Aspects*. Academic, New York; 1978.
25. Brockhurst MA, Harrison E, Hall JPJ, Richards T, McNally A, MacLean C: *The Ecology and Evolution of Pangenomes*. Cell Press; 2019.
26. •• Horesh G, Taylor-Brown A, McGimpsey S, Lassalle F, Corander J, Heinz E *et al.*: **Different evolutionary trends form the twilight zone of the bacterial pan-genome**. *Microb Genomics* 2021, **7**:000670

Describe a novel classification of core and accessory genes based on their frequency distribution across lineages.

27. Koonin EV, Wolf YI: **Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world**. *Nucleic Acids Res* 2008, **36**:6688-6719.
28. Zong Z, Fenn S, Connor C, Feng Y, McNally A: **Complete genomic characterization of two *Escherichia coli* lineages responsible for a cluster of carbapenemresistant infections in a Chinese hospital**. *J Antimicrob Chemother* 2018, **9**:2340-2346.
29. Matamoros S, Van Hattem JM, Arcilla MS, Willemse N, Melles DC, Penders J *et al.*: **Global phylogenetic analysis of *Escherichia coli* and plasmids carrying the mcr-1 gene indicates bacterial diversity but plasmid restriction**. *Sci Rep* 2017:15364.
30. Oteo J, Diestra K, Juan C, Bautista V, Novais Pérez-Vázquez M *et al.*: **Extended-spectrum β-lactamase-producing *Escherichia coli* in Spain belong to a large variety of multilocus sequence typing types, including ST10 complex/A, ST23 complex/A and ST131/B2**. *Int J Antimicrob Agents* 2009, **34**:173-176.
31. Bosi E, Fondi M, Orlandini V, Perrin E, Maida I, de Pascale D *et al.*: **The pangenome of (Antarctic) *Pseudoalteromonas* bacteria: evolutionary and functional insights**. *BMC Genomics* 2017, **18**:93.
32. Spring-Pearson SM, Stone JK, Doyle A, Allender CJ, Okinaka RT, Mayo M *et al.*: **Pangenome analysis of *Burkholderia pseudomallei*: genome evolution preserves gene order despite high recombination rates**. *PLoS One* 2015, **10**: e0140274.
33. Hall RJ, Whelan FJ, Cummins EA, Connor C, McNally A, McInerney JO: **Gene-gene relationships in an *Escherichia coli* accessory genome are linked to function and mobility**. *Microb Genomics* 2021, **7**:000650.
34. Nakamura K, Murase K, Sato MP, Toyoda A, Itoh T, Mainil JG *et al.*: **Differential dynamics and impacts of prophages and plasmids on the pangenome and virulence factor repertoires of Shiga toxin-producing *Escherichia coli* O145:H28**. *Microb Genomics* 2020, **6**:e000323.
35. Wyres KL, Wick RR, Judd LM, Froumine R, Tokolyi A, Gorrie CL *et al.*: **Distinct evolutionary dynamics of horizontal gene transfer in drug resistant and virulent clones of *Klebsiella pneumoniae***. *PLoS Genet* 2019, **15**:e1008114.

36. • Moulana A, Anderson RE, Fortunato CS, Huber JA: **Selection is a significant driver of gene gain and loss in the pangenome of the bacterial genus *Sulfurovum* in geographically distinct deep-sea hydrothermal vents**. *mSystems* 2020, **5** 00673–19
Role of eco-evolutionary dynamics, specifically metabolic variations, in shaping pangenomes.

37. Argemi X, Matelska D, Ginalski K, Riegel P, Hansmann Y, Bloom J *et al.*: **Comparative genomic analysis of *Staphylococcus lugdunensis* shows a closed pan-genome and multiple barriers to horizontal gene transfer**. *BMC Genomics* 2018, **19**:621.

38. Wheatley RM, MacLean RC: **CRISPR-cas systems restrict horizontal gene transfer in *Pseudomonas aeruginosa***. *ISME J* 2021, **15**:1420-1433.

39. Eppinger M, Worsham PL, Nikolich MP, Riley DR, Sebastian Y, Mou S *et al.*: **Genome sequence of the deep-rooted *Yersinia pestis* strain Angola reveals new insights into the evolution and pangenome of the plague bacterium**. *J Bacteriol* 2010, **192**:1685-1699.

40. McCutcheon JP, Moran NA: **Extreme genome reduction in symbiotic bacteria**. *Nat Rev Microbiol* 2011, **10**:13-26.

41. Tamas I, Klasson L, Canback B, Naslund AK, Eriksson AS, Wernegreen JJ *et al.*: **50 million years of genomic stasis in endosymbiotic bacteria**. *Science* 2002, **296**:2376-2379.

42. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ *et al.*: **Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen**. *Nature* 2000, **406**:959-964.

43. Wolfgang MC, Kulasekara BR, Liang X, Boyd D, Wu K, Yang Q *et al.*: **Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa***. *Proc Natl Acad Sci U S A* 2003, **100**:8484-8489.

44. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM *et al.*: **Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain**. *Microbiology* 2012, **4**:1005-1015.

45. Evans R, Beckerman AP, Wright RCT, McQueen-Mason S, Bruce NC, Brockhurst MA: **Eco-evolutionary dynamics set the tempo and trajectory of metabolic evolution in multispecies communities**. *Curr Biol* 2020, **30**:4984-4988.