Metabolic gene clusters, fungal diversity, and the generation of accessory functions

Jason C. Slot[1]* & Emile Gluck-Thaler[1#]

[1]Department of Plant Pathology, The Ohio State University, Columbus, OH, USA

[*]Corresponding author: slot.1@osu.edu; 481C Kottman Hall, 2021 Coffey Road, Columbus, OH, USA; 1 (614) 688-2122

[#]gluckthaler.1@osu.edu

Keywords: supergene; secondary metabolism; biosynthesis; degradation; genome architecture

Short title: Gene clusters and the generation of accessory functions

Abstract

Ecological interactions are largely determined by adaptive traits deemed "accessory". In plants, fungi, and bacteria, such traits mainly comprise metabolic pathways that produce or transform diverse molecules. While accessory metabolic pathways are pervasive, it is often difficult to identify their genetic bases. Recently, in-depth descriptions of metabolic gene clusters (MGCs), which encode discrete metabolic pathways, have greatly simplified the characterization of genotype-phenotype maps, yet questions of how this genome architecture relates to the evolution of accessory functions remain. Fungi are uniquely positioned to spearhead investigations into these dynamics because they display gradients in clustering across pathways and taxa. This review will focus on the role of MGCs as both agents and consequences of the accessory function evolution that underpins fungal diversification.

MGCs: repositories of accessory metabolic functions

Accessory metabolic functions underpin ecological adaptation and innovation in plants, fungi, and bacteria by determining the outcomes of organismal interactions [1]. Chemical defenses, for example, enable these organisms to overcome competition and resist predation despite their limited mobility. Nutritionally adaptive metabolic pathways can facilitate the breakdown of challenging substrates or protect digested resources from competitors. The accessory metabolites of complex organisms further define the adaptive microbial communities with which they associate [2]. However, there is a general lack of knowledge about the individual genes responsible for these phenotypes. This problem has been partially addressed by investigations at the -omics scale, e.g. inferring and describing networks of accessory genes whose expression or distributions are associated with ecologically important phenotypes [3,4].

Genome structure provides an alternative window into accessory metabolic functions. Specifically, accessory metabolisms are frequently encoded by Metabolic Gene Clusters (MGCs) composed of neighboring, coordinately-expressed genes participating in discrete metabolic pathways, often accompanied by supportive functions like regulation, transport, and self-protection from the encoded phenotype [5, 6]. Because genes cluster in response to natural selection, MGCs are effective proxies of adaptive metabolic functions, and can therefore greatly accelerate the discovery and characterization of accessory pathways contributing to ecological phenotypes [7]. Convergent origins and interspecies exchanges of MGCs can also indicate selection on the accessory functions they encode [8, 9, 10]. Further, the ecology and evolution of MGCs can be inferred independently of the genetic mechanisms and traits they encode, enabling large-scale modeling of chemo-diversification processes [11].

One way to identify accessory function-encoding MGCs is by conducting targeted searches using probabilistic models trained on previously identified clusters. AntiSMASH and the associated ClusterFinder module are general-purpose tools for mining biosynthetic MGCs by similarities in gene sequence, cluster composition, and the spatial distributions of protein family domains in genomes [11, 12]. These tools can be used to identify the genetic basis of real and predicted natural products [13] and further compare potential chemodiversity across fungal populations and lineages [14, 15]. Because such programs use models trained on known pathways/MGCs from a limited set of fungi, bacteria, and plants, “untargeted” cluster prediction methods that use other evolutionary or ecological criteria for identifying gene clusters may be needed to uncover new types of metabolism. Some innovative untargeted approaches include searching for co-location of genes with similar promoter motifs [16], shared evolutionary histories [17], or linkage shared by unexpectedly divergent species [18, 19, 20]. The prediction of biosynthetic MGCs has been the major focus of genome mining initiatives to date, though catabolic MGC discovery is poised to follow suit [18].

MGCs interact with the drivers and constraints of accessory function evolution

Genes for accessory functions can arise through vertical gene duplication (VGD-duplication of a locus within a genome, inherited by offspring), horizontal gene transfer (HGT-duplication from a locus in one genome to a locus in another)(Figure 1), or *de novo* gene evolution [21]. Genes that already perform accessory roles, such as dedicated secondary metabolism structural genes (Nonribosomal Peptide Synthetases, Polyketide Synthetases, etc.) may be the main source of novel accessory genes. Truly novel metabolism also emerges

through duplication of core metabolic genes [22, 23], but it is difficult to distinguish such accessory genes from the core genes from which they are derived purely based on amino acid sequence. MGCs are hotspots for VGD and HGT [24]. These closely-related information copying processes remove the evolutionary constraints that were imposed on the original genes by genetic networks under natural selection; they facilitate adaptive evolution by allowing partitioning of ancestral functions (subfunctionalization) or the emergence of novel functions [25, 26, 27]. HGT of gene modules and MGCs may further impact organismal fitness by copying environmentally adaptive functions into novel genetic backgrounds [9, 28, 29]. A propensity for either type of gene duplication can increase an organism's latent capacity to evolve and elaborate on accessory functions [30]. HGT of MGCs and multigene modules they contain may further accelerate adaptation by bypassing fitness valleys (e.g. toxic transitional states) encountered during stepwise addition of genes to a genome [5, 31].

Linkage, specialization, and the generation of adaptive accessory functions

MGCs provide a framework for testing outstanding hypotheses in evolutionary biology concerning the origins of adaptive phenotypes. For example, enzyme activity by default is promiscuous, which furnishes important variation for evolution of accessory functions, but carries the burden of pathway inefficiency and the potential for toxic interference among pathways [32]. Specialization, in contrast, entails functional optimization and metabolic efficiency, but decreases the overall evolutionary potential of the specialized genes [32]. MGCs may mediate some of these trade-offs by enabling coordination of enzyme function and regulation, reducing their interference with other metabolic processes, while still maintaining promiscuous functions [8]. By decreasing the fitness costs of functionally linked accessory genes, clustering preserves adaptive gene combinations and facilitates their joint incorporation into new phenotypic contexts, accelerating the emergence of accessory function [33]. Differential rates of degeneration and loss among protein functional classes within MGCs lend further insight into the process of adaptive specialization. Recent studies have demonstrated that enzymes are preferentially lost from MGCs while regulators and transporters are selectively retained [34, 35], suggesting specialization could result in enzymatic "dead ends", and that continued innovation requires a regular replenishment of enzymes from conserved or promiscuous functions. The ability to generate and preserve functional linkages in gene clusters would thus impact an organism's ability to generate and incorporate accessory diversity (i.e. its evolvability)[36].

Variation in clustering provides insight into the mechanisms generating accessory functions

Fungi are uniquely positioned to elucidate the mechanisms generating accessory functions and MGC diversity because their known genomic diversity captures variation in clustering across metabolic pathways and across lineages in ways that bacteria, with highly clustered genomes, and plants, with few clusters in their genomes, do not (Figure 1). Specifically, there are significant differences in clustering rates between the two filamentous fungal subphyla, Pezizomycotina and Agaricomycotina. While hundreds of MGCs have been identified in Pezizomycotina, there are only five Agaricomycotina pathways deposited in the MIBiG database v1.4 (mibig.secondarymetabolites.org/repository.html), and only a few more can be found in the literature (Table 1). The MGCs identified in Pezizomycotina fungi further

differ in compositional complexity and the metabolism encoded compared to those identified in Agaricomycotina (Table 1). Known or predicted Agaricomycotina clusters are small, and contain recently duplicated genes, compared to those in Pezizomycotina, which cover a large range of sizes and are compositionally more diverse [37]. Additionally, Agaricomycete biosynthetic pathways employ parallel processes at specific metabolic steps and in the classes of compounds produced [38]. These distinctions suggest accessory metabolism and MGCs in Agaricomycotina have commonalities with those of plants [39], while the genomic structure of accessory metabolism in Pezizomycotina is more similar to bacteria [11]. This overlap complicates assumptions about cluster evolution, where tandem duplications provide a neutral evolutionary mechanism for cluster formation in eukaryotes, and the tendency for genes to be transcribed in operons offers a regulatory mechanism selecting for clustering in bacteria. By providing a range of compositional complexity, but with limited operon-like transcription, the spectrum of fungal gene clustering forces attention to other evolutionary mechanisms driving gene clustering.

To illustrate these trends (Figure 1), certain cluster-encoded metabolisms, such as those based on highly modular nonribosomal peptide synthetase enzymes, along with numerous supporting enzymes, transporters and regulators, are largely restricted to morphologically simple fungi and bacteria [37, 40]. For example, Pezizomycotina have an average of ~10 NRPS containing clusters per genome versus 1.2 in Agaricomycotina. Interestingly, mushroom-forming fungi still produce non-protein peptide metabolites with ribosomally generated oligopeptides encoded in high copy number MSDIN genes, whose products diversify through simple coding mutations [41]. Terpenoid biosynthesis, where biologically meaningful diversity can be readily generated through small modifications and high enzyme promiscuity, so far also seems to be over-represented in morphologically complex fungi and plants [37, 42]. The difference in genome architecture among fungi is especially pronounced in phenolic degradation. While putative phenolic degradation clusters are diverse and complex across Pezizomycotina [18], in Agaricomycotina, most such clustering is the result of tandem duplications and few putative clusters contain more than a single gene family [43]. Disentangling the roles of species-specific metabolism and lineage-based differences in genome evolution mechanisms will become more feasible with greater genome sampling and ecological knowledge of fungi.

Life history traits and genome architecture theory predict a clustering gradient across fungi

It is not entirely clear whether the absolute numbers of clusters identified in the respective lineages reflect differences in propensity to cluster, or if different cluster compositions in Agaricomycotina leave them less likely to be detected using biased models and methods. Nonetheless, the variation we observe is predicted from complex interactions between migration, recombination rates, and selection [44], which differ between fungal subphyla as a result of contrasting reproductive strategies, generation times, and ecological lifestyles. Under strong selection and sufficient gene flow, clustered architectures are favored, because they prevent the influx of maladaptive alleles and preserve the fitness of locally adapted genotypes [44]. For example, when genes participate in the same specialized pathway, clusters reduce recombination between differently optimized enzymes that would be detrimental to their cooperation in metabolic transformations.

The mating strategies of individual organisms may strongly influence genome architecture, the rates of VGD and HGT, and the diversity of accessory functions (Figure 1). For example, many bacteria and fungi participate in large "pan-genomes" containing far greater diversity than can be contained in a single individual. Sharing in pan-genomes is facilitated by mobile elements and unequal recombination mechanisms, and gives access to diverse gene families, gene combinations, and sequence variants. In these types of populations, clusters are more prevalent and they enable the parallel diversification of accessory functions across numerous genomes under diverse environmental selection. In this way, pan-genomes result in varied assortments of discrete pathways in individual genomes. By contrast, fungi and plants that participate in smaller pan-genomes, will have more limited access to gene variants and variable gene combinations, resulting in a smaller number of highly branched and parallel pathways within individual lineages. Pathways in these organisms are constrained to evolve primarily through VGD and intragenomic mutation. Differences in the types of metabolic diversity and in the composition of MGCs found across bacteria, fungi, and plants are consistent with these processes, but specific tests remain to be done.

Ecology is similarly seen as a driver of reproductive strategies that impact selection on genome architecture (Figure 2). Tree-decaying Agaricomycotina often have large, long-lived somatic biomasses with rare, periodic meiotic sexual reproduction. In contrast, the typically small units of rapidly sporulating Pezizomycotina commonly engage in clonal reproduction and self-fertilization, in addition to sexual reproduction. Pervasive, non-meiotic recombination (including parasexuality) in Pezizomycotina [45], similar to bacteria, may explain their substantial pan-genomes and elevated gene flow between populations and across species barriers. The increased proclivity for parasexuality may also favor non-homologous recombination and structural rearrangements of large linkage blocks (even entire chromosomes) that repress recombination. Repression of recombination frees nascent gene clusters to accumulate transposed genes and to drift while becoming selected as a block, effectively acting as a superallele in a population's pan-genome [46, 47]. Indeed, a number of recent studies have confirmed the existence of "idiomorphic" MGC loci in Pezizomycotina fungi, where variably distributed MGCs occupy the same genomic region in different individuals [14, 48, 49, 50, 51]. Supernumerary (dispensible) chromosomes are an extreme example of parasexual/horizontal inheritance of co-adapted blocks of accessory functions. HGT of supernumerary chromosomes in *Fusarium* has been shown to enable infection of a new host by confering multiple adaptive functions [52]. Decreased propensity for genomic rearrangements and constraints on gene flow in Agariciomycotina due to sexual recombination may relax the drive to cluster in this lineage compared with Pezizomycotina, as decreased population-level rates of small-scale genomic rearrangements are also associated with decreased rates of clustering [53]. Intriguingly, ecological selection for self-mating appears to drive formation of mating locus clusters, which relaxes some evolutionary constraints imposed on predominantly sexual populations [54, 55].

One line of evidence supporting the hypothesis that the spectrum of lifestyles and reproductive strategies of fungi impacts patterns of metabolic diversification is the rate of MGC horizontal duplication. It is notable that to date, while large HGTs of MGCs are more commonly inferred in Pezizomycotina (commensurate with higher HGT in general [24]), the only published case of MGC HGT in Agaricomycotina involves dung-decay mushrooms. Dung fungi are driven to

inbreed and reproduce more rapidly due to the patchiness and ephemeral nature of this resource [56]. The case of HGT of the psilocybin cluster among dung-decay fungi [10], contrasts with the other evolutionarily well-characterized Agaricomycotina MGC for luciferin, which evolves through concerted VGD and loss [57]. However, these hypotheses need to be rigorously tested.

Competition among osmotrophs is thought to be the main driver of the diversification of accessory pathways [58]. Interestingly the genomes of fungus-like oomycetes (stramenopiles) are structured more similarly to plants, with extensive tandem gene duplication, but few gene clusters and little accessory metabolism [59, 60]. While oomycetes are similar to fungi in physiology, and their range of ecological roles, they differ by having a dominant diploid stage of their life-cycles, suggesting reproductive strategies as well as nutritional mode are important in genome architecture.

A single gene family illustrates different evolutionary processes among fungal lineages

Different patterns of accessory enzyme diversification in Pezizomycotina and Agaricomycotina exemplify the interplay between accessory functions and gene clustering. For example, phylogenetic analysis supports the emergence of single-domain MT-33 N-methyltransferases by partial duplication of bifunctional ergothioneine synthases in both lineages [61]. These enzymes alternately participate in ergotamine toxin (Pezizomycotina) and methylated tryptophan biosynthesis (Agaricomycotina). Across Agaricomycotina, there have been many recent origins of these enzymes by VGD, and the only observed ancient paralog has itself been subject to ongoing loss. Furthermore, there are no indications that these accessory paralogs are in MGCs or horizontally transferred in Agaricomycotina. In Pezizomycotina, by contrast, while these N-methyltransferases perform a similar function N-methylating indole alkaloids, they have rarely emerged by VGD, and they appear to disperse by HGT as part of complex MGCs [61, 62]. In short, these novel enzymes appear constrained to evolve simple peripheral functions through VGD in isolated lineages of Agaricomycotina, but participate in complex clustered pathways that are shared across species through HGT in Pezizomycotina.

MGC distributions imply ecological functions

The ecological impact of accessory functions in microorganisms can be difficult to discern because species may be opportunistic on the substrate of their isolation, and the genes encoding these functions are often unknown [63]. Equally confounding, sub-populations of a species may be specialized for alternative cryptic niches. MGCs can help disentangle the importance of specific functions for ecological fitness because they enable the rapid identification of genes and gene families of interest contributing to specific phenotypes. For example, the biased distribution and HGT of clustered genes involved in psilocybin production among dung- and insect- associated fungi, coupled with the known effects of psilocybin on serotonin-receptors, led to speculation that this tryptamine alkaloid reduces mycophagy by insects by interfering with neural signaling. Subsequent detection of psilocybin (by a likely a convergent psilocybin pathway) in cicada parasitizing fungi provided additional evidence that this compound is selected for its effects on insects [64]. Similarly, the identification of a MGC found previously in pathogens of woody plants enabled a reverse ecology approach to discovering new hosts for several grass pathogens [7]. And finally, targeted and untargeted

searches for MGCs have further revealed previously unexplored dimensions of accessory function evolution and led to the proposal of their ecological associations [18].

Summary

Accessory functions are the essence of ecological adaptation and innovation, yet they often resist characterization. Genome architecture-focused approaches, such as MGC mining and characterization, promise a relatively straightforward path to cut through genomic noise and decipher ecological relationships in parallel with molecular functions [63, 65]. Although MGCs are observed across all major Kingdoms of life, the diversity of fungal mating strategies and lifestyles is likely responsible for generating fungal-specific variation in clustering rates, positioning fungi and their MGCs as model systems for linking specific ecological, demographic, and genetic processes with the generation of accessory metabolic functions (Figure 2). In particular, elucidating the relative impact of vertical and horizontal gene family expansion (VGD and HGT) on different strategies for metabolic diversification should greatly improve our ability to explain how, when, and why accessory functions evolve.

Figure Legends

Figure 1: *The spectrum of accessory metabolic compartmentalization and diversification.* Bacteria, Ascomycetes, Basidiomycetes, and Plants (shown as silhouettes) capture gradients in accessory metabolic diversification and the storage of accessory metabolic functions within and across individuals. Moving from left to right, pan-genomes become progressively closed (i.e., less variable among individuals), genomic architectures feature fewer metabolic gene clusters encoding accessory functions, and individual genomes transition from encoding many different specialized accessory pathways to few highly diversified pathways. The major candidate mechanisms contributing to the gradients are shown below, and are expected to manifest to different degrees in each of the depicted lineages.

Figure 2: *The eco-evolutionary dynamics of genome architecture*. Ecological selection is the overarching force shaping the evolution of accessory functions, acting through evolutionary processes operating at multiple scales, as discussed in the text. Ecological selection pressures determine the most adaptive population dynamics, including mating/reproduction strategies, optimal effective population size (N_E), and migration rates. In turn, the spectra of these population dynamics determine the mechanisms of genome evolution and metabolic diversification that are ultimately available for generating adaptive variation. Recombination mechanisms and streamlining vary across different life-history strategies and result in the types of evolvability favored and the size of pan-genomes (Figure 1). These drivers simultaneously favor different modes of metabolic diversification such as the combination of gene functions in novel pathways, and their differential optimization/specialization. Contrasting patterns of genome architecture (e.g., gene clustering vs. gene family expansion) are tangible outcomes of interactions among these processes (Figure 1), but they may also feed back on and influence the impact of the main evolutionary drivers. For example, gene clustering favors HGT and pathway specialization, which can mitigate competition and increase access to resources.

Acknowledgements

This work was supported by the National Science Foundation (DEB-1638999, JCS) and the Fonds de Recherche du Québec-Nature et Technologies (EG-T).

The authors declare no conflict of interest.

References

[1] Raguso, Robert A., et al. "The raison d'être of chemical ecology." Ecology 96.3 (2015): 617-630.

[2] Huang AC, Jiang T, Liu YX, Bai YC, Reed J, Qu B, Goossens A, Nützmann HW, Bai Y, Osbourn A. A specialized metabolic network selectively modulates Arabidopsis root microbiota. Science. 2019 May 10;364(6440):eaau6389. ••This comprehensive study links plant MGC function and diversity with functional ecological consequences.

[3] Cary JW, Entwistle S, Satterlee T, Mack BM, Gilbert MK, Chang PK, Scharfenstein L, Yin Y, Calvo AM. The Transcriptional Regulator Hbx1 Affects the Expression of Thousands of Genes in the Aflatoxin-Producing Fungus Aspergillus flavus. G3: Genes, Genomes, Genetics. 2019 Jan 1;9(1):167-78.

[4] Nielsen JC, Prigent S, Grijseels S, Workman M, Ji B, Nielsen J. Comparative Transcriptome Analysis Shows Conserved Metabolic Regulation during Production of Secondary Metabolites in Filamentous Fungi. Msystems. 2019 Apr 30;4(2):e00012-19.

[5] Keller NP. Translating biosynthetic gene clusters into fungal armor and weaponry. Nature chemical biology. 2015 Sep;11(9):671.

[6] Kjærbølling I, Vesth T, Andersen MR. Resistance Gene-Directed Genome Mining of 50 Aspergillus species. mSystems. 2019 Aug 27;4(4):e00085-19. ••This manuscript combines guilt-by-association gene cluster detection with the concept that MGCs encode self-protection mechanisms.

[7] Thynne E, Mead OL, Chooi YH, McDonald MC, Solomon PS. Acquisition and loss of secondary metabolites shaped the evolutionary path of three emerging phytopathogens of wheat. Genome biology and evolution. 2019 Feb 21;11(3):890-905.

[8] Gluck-Thaler E, Vijayakumar V, Slot JC. Fungal adaptation to plant defences through convergent assembly of metabolic modules. Molecular ecology. 2018 Dec;27(24):5120-36.

[9] Ocaña-Pallarès E, Najle SR, Scazzocchio C, Ruiz-Trillo I. Reticulate evolution in eukaryotes: Origin and evolution of the nitrate assimilation pathway. PLoS genetics. 2019 Feb 21;15(2):e1007986. •This study revisits the first convincing case of complete fungal gene cluster HGT, the nitrate assimilation cluster, and finds there was a more complex history of transfers involving additional, distant eukaryotic lineages.

[10] Reynolds HT, Vijayakumar V, Gluck-Thaler E, Korotkin HB, Matheny PB, Slot JC. Horizontal gene cluster transfer increased hallucinogenic mushroom diversity. Evolution Letters. 2018 Apr;2(2):88-101.

[11] Cimermancic P, Medema MH, Claesen J, Kurita K, Brown LC, Mavrommatis K, Pati A, Godfrey PA, Koehrsen M, Clardy J, Birren BW. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell. 2014 Jul 17;158(2):412-21.

[12] Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic acids research. 2019 Apr 29.

[13] Theobald S, Vesth TC, Rendsvig JK, Nielsen KF, Riley R, de Abreu LM, Salamov A, Frisvad JC, Larsen TO, Andersen MR, Hoof JB. Uncovering secondary metabolite evolution and

biosynthesis using gene cluster networks and genetic dereplication. Scientific reports. 2018 Dec 18;8(1):17957.
[14] Lind AL, Wisecaver JH, Lameiras C, Wiemann P, Palmer JM, Keller NP, Rodrigues F, Goldman GH, Rokas A. Drivers of genetic diversity in secondary metabolic gene clusters within a fungal species. PLoS biology. 2017 Nov 17;15(11):e2003583.
[15] Olarte RA, Menke J, Zhang Y, Sullivan S, Slot JC, Huang Y, Badalamenti JP, Quandt AC, Spatafora JW, Bushley KE. Chromosome rearrangements shape the diversification of secondary metabolism in the cyclosporin producing fungus Tolypocladium inflatum. BMC genomics. 2019 Dec;20(1):120.
[16] Wolf T, Shelest V, Nath N, Shelest E. CASSIS and SMIPS: promoter-based prediction of secondary metabolite gene clusters in eukaryotic genomes. Bioinformatics. 2015 Dec 9;32(8):1138-43.
[17] Slot JC. Fungal gene cluster diversity and evolution. In Advances in genetics 2017 Jan 1 (Vol. 100, pp. 141-178). Academic Press.
[18] Gluck-Thaler E, Slot JC. Specialized plant biochemistry drives gene clustering in fungi. The ISME journal. 2018 Jul;12(7):1694. ••This study systematically reveals a wealth of previously undescribed catabolic MGCs in fungal genomes through untargeted MGC prediction.
[19] Winter S, Jahn K, Wehner S, Kuchenbecker L, Marz M, Stoye J, Böcker S. Finding approximate gene clusters with Gecko 3. Nucleic acids research. 2016 Sep 26;44(20):9600-10.
[20] de Jonge R, Ebert MK, Huitt-Roehl CR, Pal P, Suttle JC, Spanner RE, Neubauer JD, Jurick WM, Stott KA, Secor GA, Thomma BP. Gene cluster conservation provides insight into cercosporin biosynthesis and extends production to the genus Colletotrichum. Proceedings of the National Academy of Sciences. 2018 Jun 12;115(24):E5459-66. •This study demonstrates the utility of using both targeted and untargeted approaches to describing functionally-relevant MGC content by using shared synteny to help extend the boundaries of the cercosporin biosynthetic gene cluster.
[21] Vakirlis N, Hebert AS, Opulente DA, Achaz G, Hittinger CT, Fischer G, Coon JJ, Lafontaine I. A molecular portrait of de novo genes in yeasts. Molecular biology and evolution. 2017 Dec 6;35(3):631-45. ••This study develops a robust pipeline for predicting de novo genes and finds that transitions from non-coding to coding DNA occur frequently in recombination hotspots.
[22] Huang R, O'Donnell AJ, Barboline JJ, Barkman TJ. Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymes. Proceedings of the National Academy of Sciences. 2016 Sep 20;113(38):10613-8.
[23] Copley SD. Evolution of a metabolic pathway for degradation of a toxic xenobiotic: the patchwork approach. Trends in biochemical sciences. 2000 Jun 1;25(6):261-5.
[24] Wisecaver JH, Slot JC, Rokas A. The evolution of fungal metabolic pathways. PLoS genetics. 2014 Dec 4;10(12):e1004816.
[25] Voordeckers K, Brown CA, Vanneste K, van der Zande E, Voet A, Maere S, Verstrepen KJ. Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. PLoS biology. 2012 Dec 11;10(12):e1001446.

[26] Lynch M, O'Hely M, Walsh B, Force A. The probability of preservation of a newly arisen gene duplicate. Genetics. 2001 Dec 1;159(4):1789-804.
[27] Wapinski I, Pfeffer A, Friedman N, Regev A. Natural history and evolutionary principles of gene duplication in fungi. Nature. 2007 Sep;449(7158):54.
[28] Kominek J, Doering DT, Opulente DA, Shen XX, Zhou X, DeVirgilio J, Hulfachor AB, Groenewald M, Mcgee MA, Karlen SD, Kurtzman CP. Eukaryotic acquisition of a bacterial operon. Cell. 2019 Mar 7;176(6):1356-66. ••Although early speculated as an explanation for fungal gene clusters, this is the first convincing example of whole operon HGT from bacteria to fungi.
[29] Milner DS, Attah V, Cook E, Maguire F, Savory FR, Morrison M, Müller CA, Foster PG, Talbot NJ, Leonard G, Richards TA. Environment-dependent fitness gains can be driven by horizontal gene transfer of transporter-encoding genes. Proceedings of the National Academy of Sciences. 2019 Mar 19;116(12):5613-22. ••The consequences of HGT are often unknown. This study is one of the few to demonstrate how acquiring novel genes can alter organismal fitness.
[30] Wagner A. Gene duplications, robustness and evolutionary innovations. Bioessays. 2008 Apr;30(4):367-73.
[31] Lawrence JG, Roth JR. Selfish operons: horizontal transfer may drive the evolution of gene clusters. Genetics. 1996 Aug 1;143(4):1843-6
[32] Copley SD. Shining a light on enzyme promiscuity. Current opinion in structural biology. 2017 Dec 1;47:167-75.
[33] Del Carratore F, Zych K, Cummings M, Takano E, Medema MH, Breitling R. Computational identification of co-evolving multi-gene modules in microbial biosynthetic gene clusters. Communications Biology. 2019 Feb 28;2(1):83. ••This study is the first to develop tools for specifically identifying conserved gene modules within MGCs, setting the stage for advancing our understanding of how modular, combinatorial processes shape MGC evolution.
[34] Reynolds HT, Slot JC, Divon HH, Lysøe E, Proctor RH, Brown DW. Differential retention of gene functions in a secondary metabolite cluster. Molecular biology and evolution. 2017 Apr 28;34(8):2002-15. •This study points to a pattern of selective retention of the supporting genes (regulators and transporters) of metabolic gene clusters that are highly prone to loss.
[35] Campbell MA, Staats M, van Kan JA, Rokas A, Slot JC. Repeated loss of an anciently horizontally transferred gene cluster in Botrytis. Mycologia. 2013 Sep 1;105(5):1126-34.
[36] Pepper JW. The evolution of evolvability in genetic linkage patterns. Biosystems. 2003 May 1;69(2-3):115-26.
[37] Rokas A, Wisecaver JH, Lind AL. The birth, evolution and death of metabolic gene clusters in fungi. Nature Reviews Microbiology. 2018 Sep 7:1.
[38] Wick J, Heine D, Lackner G, Misiek M, Tauber J, Jagusch H, Hertweck C, Hoffmeister D. A fivefold parallelized biosynthetic process secures chlorination of Armillaria mellea (honey mushroom) toxins. Appl. Environ. Microbiol.. 2016 Feb 15;82(4):1196-204.
[39] Wisecaver JH, Borowsky AT, Tzin V, Jander G, Kliebenstein DJ, Rokas A. A global coexpression network approach for connecting genes to specialized metabolic pathways in plants. The Plant Cell. 2017 May 1;29(5):944-59. •• The authors of this

study found that few networks of co-expressed genes correspond to predicted MGCs in plant genomes, suggesting rates of clustering within this lineage are even lower than previously thought.
[40] Medema MH, Cimermancic P, Sali A, Takano E, Fischbach MA. A systematic computational analysis of biosynthetic gene cluster evolution: lessons for engineering biosynthesis. PLoS computational biology. 2014 Dec 4;10(12):e1004016.
[41] Luo H, Hong SY, Sgambelluri RM, Angelos E, Li X, Walton JD. Peptide macrocyclization catalyzed by a prolyl oligopeptidase involved in α-amanitin biosynthesis. Chemistry & biology. 2014 Dec 18;21(12):1610-7.
[42] Quin MB, Flynn CM, Schmidt-Dannert C. Traversing the fungal terpenome. Natural product reports. 2014;31(10):1449-73.
[43] Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martínez AT, Otillar R, Spatafora JW, Yadav JS, Aerts A. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. Science. 2012 Jun 29;336(6089):1715-9.
[44] Yeaman S, Whitlock MC. The genetic architecture of adaptation under migration–selection balance. Evolution: International Journal of Organic Evolution. 2011 Jul;65(7):1897-911.
[45] Barcellos FG, Hungria M, Pizzirani-Kleiner AA. Limited vegetative compatibility as a cause of somatic recombination in Trichoderma pseudokoningii. Brazilian Journal of Microbiology. 2011 Dec;42(4):1625-37.
[46] Fouché S, Badet T, Oggenfuss U, Plissonneau C, Francisco CS, Croll D. Stress-driven transposable element de-repression dynamics in a fungal pathogen. bioRxiv. 2019 Jan 1:633693.
[47] Grandaubert J, Dutheil JY, Stukenbrock EH. The genomic determinants of adaptive evolution in a fungal pathogen. Evolution Letters. 2019 Jun;3(3):299-312.
[48] Gibbons JG, Salichos L, Slot JC, Rinker DC, McGary KL, King JG, Klich MA, Tabb DL, McDonald WH, Rokas A. The evolutionary imprint of domestication on genome variation and function of the filamentous fungus Aspergillus oryzae. Current Biology. 2012 Aug 7;22(15):1403-9.
[49] Plissonneau C, Hartmann FE, Croll D. Pangenome analyses of the wheat pathogen Zymoseptoria tritici reveal the structural basis of a highly plastic eukaryotic genome. BMC biology. 2018 Dec;16(1):5. •This study found that accessory genes in a highly sexual fungus tend to be organized into clusters.
[50] Uka V, Moore GG, Arroyo-Manzanares N, Nebija D, De Saeger SM, Diana Di Mavungu J. Secondary metabolite dereplication and phylogenetic analysis identify various emerging mycotoxins and reveal the high intra-species diversity in Aspergillus flavus. Frontiers in microbiology. 2019;10:667.
[51] Wollenberg RD, Sondergaard TE, Nielsen MR, Knutsson S, Pedersen TB, Westphal KR, Wimmer R, Gardiner DM, Sørensen JL. There it is! Fusarium pseudograminearum did not lose the fusaristatin gene cluster after all. Fungal biology. 2019 Jan 1;123(1):10-7.
[52] van Dam P, Fokkens L, Ayukawa Y, van der Gragt M, Ter Horst A, Brankovics B, Houterman PM, Arie T, Rep M. A mobile pathogenicity chromosome in Fusarium oxysporum for infection of multiple cucurbit species. Scientific reports. 2017 Aug 22;7(1):9042.
[53] Yeaman S. Genomic rearrangements and the evolution of clusters of locally adaptive loci. Proceedings of the National Academy of Sciences. 2013 May 7;110(19):E1743-51.

[54] Branco S, Carpentier F, de la Vega RC, Badouin H, Snirc A, Le Prieur S, Coelho MA, de Vienne DM, Hartmann FE, Begerow D, Hood ME. Multiple convergent supergene evolution events in mating-type chromosomes. Nature communications. 2018 May 21;9(1):2000.

[55] Sun S, Coelho MA, Heitman J, Nowrousian M. Convergent evolution of linked mating-type loci in basidiomycetes: an ancient fusion event that has stood the test of time. BioRxiv. 2019 Jan 1:626911.

[56] James TY. Why mushrooms have evolved to be so promiscuous: Insights from evolutionary and ecological patterns. Fungal biology reviews. 2015 Dec 1;29(3-4):167-78.

[57] Kotlobay AA, Sarkisyan KS, Mokrushina YA, Marcet-Houben M, Serebrovskaya EO, Markina NM, Somermeyer LG, Gorokhovatsky AY, Vvedensky A, Purtov KV, Petushkov VN. Genetically encodable bioluminescent system from fungi. Proceedings of the National Academy of Sciences. 2018 Dec 11;115(50):12728-32.

[58] Richards TA, Talbot NJ. Horizontal gene transfer in osmotrophs: playing with public goods. Nature Reviews Microbiology. 2013 Oct;11(10):720.

[59] McGowan J, Byrne KP, Fitzpatrick DA. Comparative Analysis of Oomycete Genome Evolution Using the Oomycete Gene Order Browser (OGOB). Genome biology and evolution. 2018 Dec 11;11(1):189-206.

[60] Judelson HS. Metabolic diversity and novelties in the oomycetes. Annual review of microbiology. 2017 Sep 8;71:21-39.

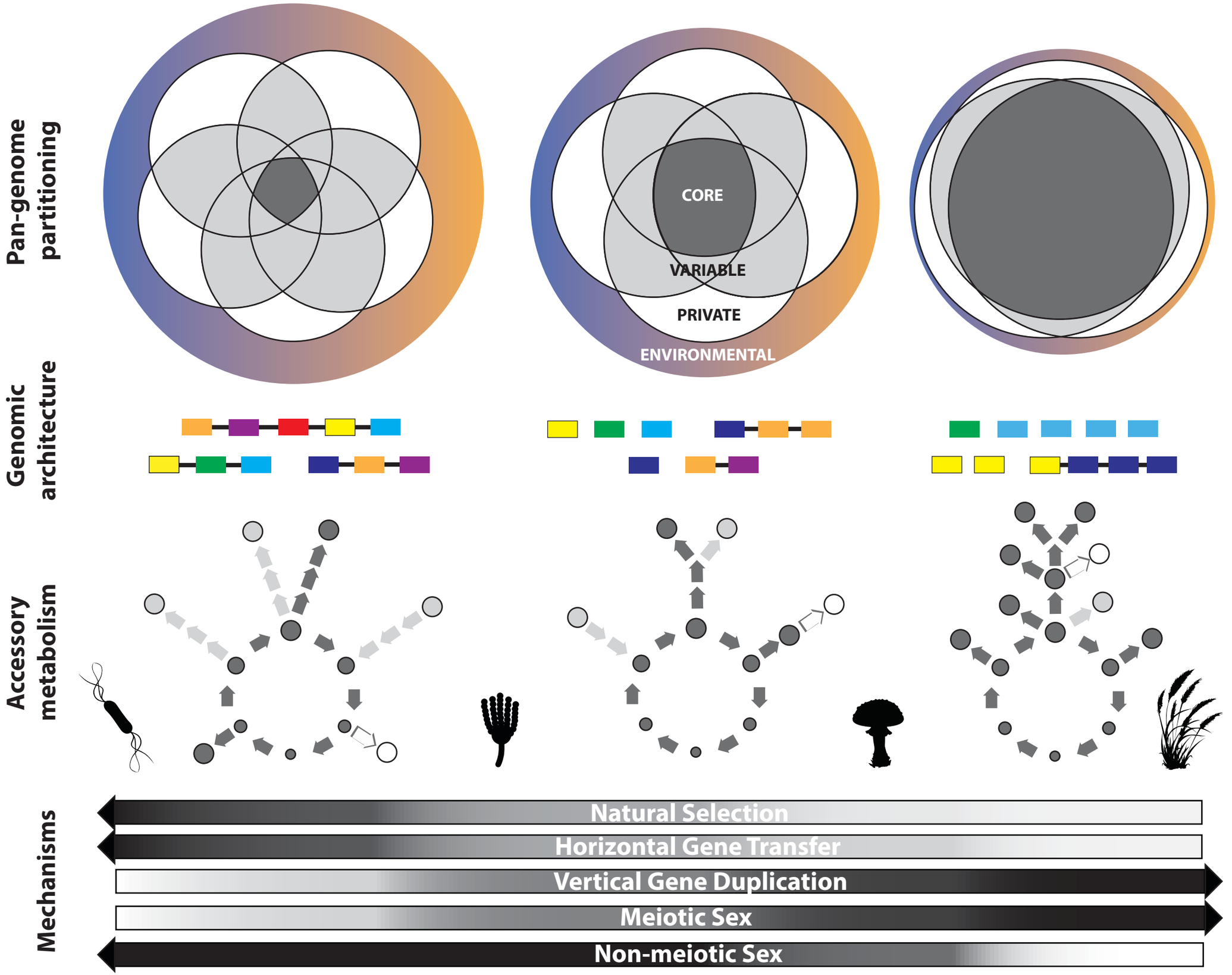
[61] Blei F, Fricke J, Wick J, Slot JC, Hoffmeister D. Iterative l-Tryptophan Methylation in Psilocybe Evolved by Subdomain Duplication. ChemBioChem. 2018 Oct 18;19(20):2160-6.

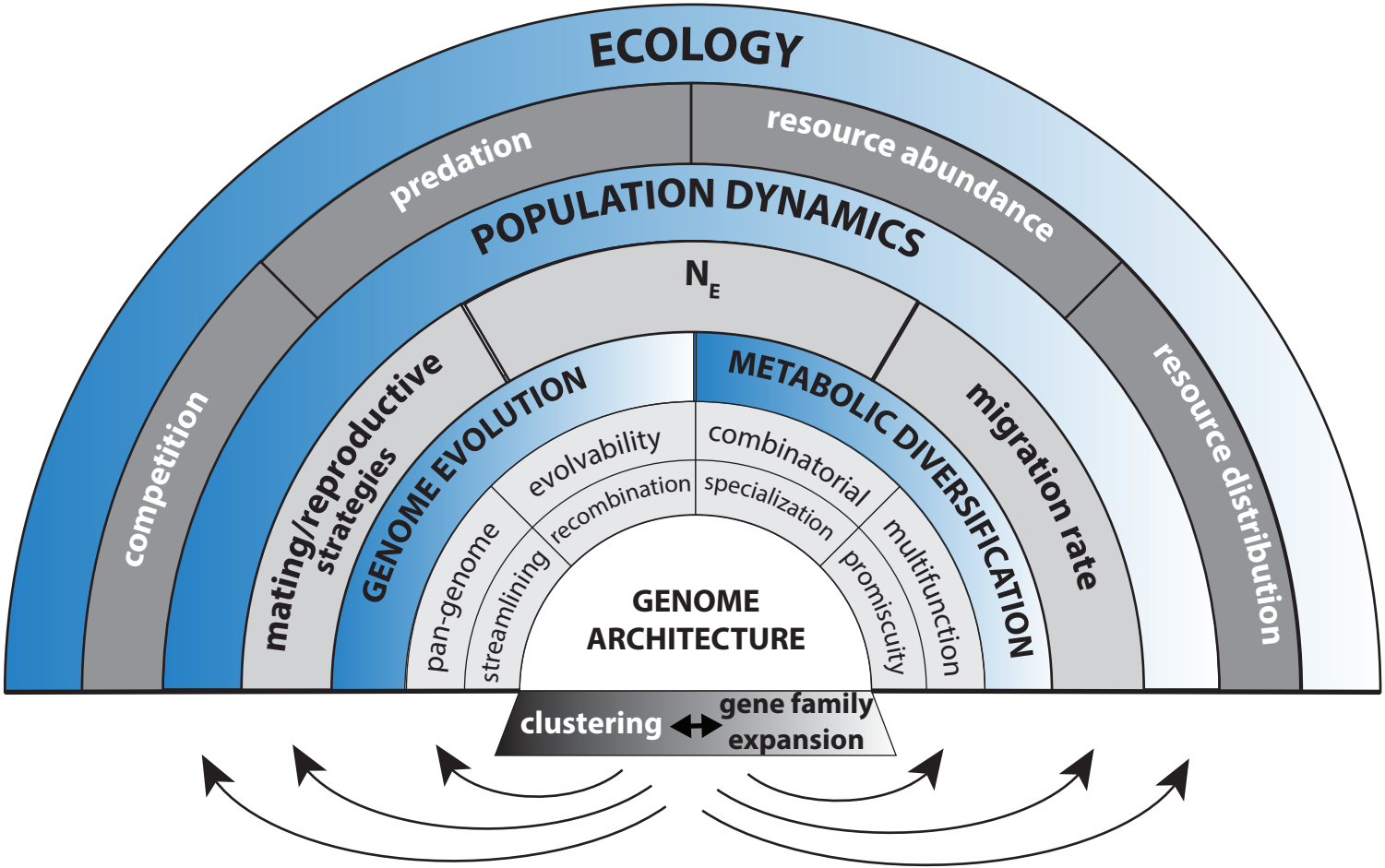
[62] Marcet-Houben M, Gabaldón T. Horizontal acquisition of toxic alkaloid synthesis in a clade of plant associated fungi. Fungal Genetics and Biology. 2016 Jan 1;86:71-80.

[63] Levy A, Gonzalez IS, Mittelviefhaus M, Clingenpeel S, Paredes SH, Miao J, Wang K, Devescovi G, Stillman K, Monteiro F, Alvarez BR. Genomic features of bacterial adaptation to plants. Nature genetics. 2018 Jan;50(1):138.

[64] Boyce GR, Gluck-Thaler E, Slot JC, Stajich JE, Davis WJ, James TY, Cooley JR, Panaccione DG, Eilenberg J, Henrik H, Macias AM. Psychoactive plant-and mushroom-associated alkaloids from two behavior modifying cicada pathogens. Fungal Ecology. 2019 Oct 1;41:147-64.

[65] Melnyk RA, Hossain SS, Haney CH. Convergent gain and loss of genomic islands drive lifestyle changes in plant-associated Pseudomonas. The ISME journal. 2019 Feb 20:1.


Pan-genome partitioning
Genomic architecture
Accessory metabolism
Mechanisms
CORE
VARIABLE
PRIVATE
ENVIRONMENTAL
Natural Selection
Horizontal Gene Transfer
Vertical Gene Duplication
Meiotic Sex
Non-meiotic Sex


ECOLOGY
predation
resource abundance
competition
resource distribution
POPULATION DYNAMICS
N_E
mating/reproductive strategies
migration rate
GENOME EVOLUTION
METABOLIC DIVERSIFICATION
evolvability
recombination
combinatorial
specialization
pan-genome
streamlining
promiscuity
multifunction
GENOME ARCHITECTURE
clustering
gene family expansion