



Protein Complexes Form a Basis for Complex Hybrid Incompatibility

Krishna B. S. Swamy^{1*}, Scott C. Schuyler^{2,3} and Jun-Yi Leu^{4*}

¹Division of Biological and Life Sciences, School of Arts and Sciences, Ahmedabad University, Ahmedabad, India,

²Department of Biomedical Sciences, College of Medicine, Chang Gung University, Taoyuan, Taiwan, ³Division of Head and Neck Surgery, Department of Otolaryngology, Chang Gung Memorial Hospital, Taoyuan, Taiwan, ⁴Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan

OPEN ACCESS

Edited by:

Igor V. Sharakhov,
Virginia Tech, United States

Reviewed by:

Katya Mack,
Stanford University, United States
Russell Corbett-Detig,
University of California, Santa Cruz,
United States

*Correspondence:

Jun-Yi Leu
jleu@imb.sinica.edu.tw
Krishna B. S. Swamy
krishna.swamy@ahduni.edu.in

Specialty section:

This article was submitted to
Evolutionary and Population
Genetics,
a section of the journal
Frontiers in Genetics

Received: 28 September 2020

Accepted: 20 January 2021

Published: 09 February 2021

Citation:

Swamy KBS, Schuyler SC and
Leu J-Y (2021) Protein Complexes
Form a Basis for Complex
Hybrid Incompatibility.
Front. Genet. 12:609766.
doi: 10.3389/fgene.2021.609766

Proteins are the workhorses of the cell and execute many of their functions by interacting with other proteins forming protein complexes. Multi-protein complexes are an admixture of subunits, change their interaction partners, and modulate their functions and cellular physiology in response to environmental changes. When two species mate, the hybrid offspring are usually inviable or sterile because of large-scale differences in the genetic makeup between the two parents causing incompatible genetic interactions. Such reciprocal-sign epistasis between inter-specific alleles is not limited to incompatible interactions between just one gene pair; and, usually involves multiple genes. Many of these multi-locus incompatibilities show visible defects, only in the presence of all the interactions, making it hard to characterize. Understanding the dynamics of protein-protein interactions (PPIs) leading to multi-protein complexes is better suited to characterize multi-locus incompatibilities, compared to studying them with traditional approaches of genetics and molecular biology. The advances in omics technologies, which includes genomics, transcriptomics, and proteomics can help achieve this end. This is especially relevant when studying non-model organisms. Here, we discuss the recent progress in the understanding of hybrid genetic incompatibility; omics technologies, and how together they have helped in characterizing protein complexes and in turn multi-locus incompatibilities. We also review advances in bioinformatic techniques suitable for this purpose and propose directions for leveraging the knowledge gained from model-organisms to identify genetic incompatibilities in non-model organisms.

Keywords: evolution, speciation, hybrid incompatibility, proteomics, proteins, bioinformatics

INTRODUCTION

Reproductive isolation impedes gene flow between species or populations and is considered fundamental to speciation (Coyne and Orr, 2004). Genomes of diverging populations accumulate differences over evolutionary time. When such populations meet to form hybrids, they may suffer from genetic incompatibilities, which are detrimental to the hybrid populations. Genetic variants between the populations that are neutral and adaptive within the populations, but deleterious between the populations are known as Dobzhansky-Muller (DM) genetic incompatibilities (Dobzhansky, 1936; Muller, 1939). This type of negative epistasis between genetic variants can cause hybrid inviability or hybrid sterility and is an important driver of

post-zygotic reproductive isolation (Presgraves, 2010a; Rieseberg and Blackman, 2010; Maheshwari and Barbash, 2011). Thus, the genes involved in DM genetic incompatibilities are commonly referred to as “speciation genes.” Although considerable effort has been devoted to identify speciation genes across several taxonomic lineages including fungi, animals, and plants, detailed molecular mechanisms underlying the failures in interactions between loci have been characterized in only a handful of cases (Wittbrodt et al., 1989; Ting et al., 1998; Brideau et al., 2006; Bomblies et al., 2007; Chang and Noor, 2007; Lee et al., 2008; Mihola et al., 2009; Brideau and Barbash, 2011; Chen et al., 2016a; Davies et al., 2016). This underlines the confounding effects of processes such as drift and linked selection, which can produce signatures similar to divergence, and also the complex genetic basis of reproductive isolation (Maheshwari and Barbash, 2011; Wolf and Ellegren, 2017). Furthermore, it should be noted that DM genetic incompatibilities are not necessarily interactions between proteins; but can also arise through the interactions of proteins with non-coding regions.

From some of these studies, there is evidence for rapid evolution in these speciation genes. Co-evolution between a rapidly evolving gene (with important function) and its partner loci are sometimes due to molecular arms races containing bouts of positive selection that can lead to the formation of speciation genes (Chou et al., 2010; Johnson, 2010; Presgraves, 2010a; Maheshwari and Barbash, 2011). One major category of such speciation genes are cyto-nuclear incompatibilities (Burton et al., 2013). Several theoretical models and experimental evidence support this class of incompatibilities across several species (Chou and Leu, 2010; Telschow et al., 2019). In yeast, almost all the known cases of DM incompatibilities are mitochondrial-nuclear incompatibilities (Lee et al., 2008; Chou et al., 2010; Chou and Leu, 2010; Presgraves, 2010a; Hou et al., 2015, 2016; Jhuang et al., 2017). Cyto-nuclear incompatibilities are also commonly observed in plants and animals (Burton et al., 2013). In recent years, genomic conflicts have been suggested to play a vital role in the formation of genetic incompatibilities (Johnson, 2010). Several known incompatibilities in plants are driven by selfish elements leading to hybrid necrosis, which has been reviewed in detail (Rieseberg and Blackman, 2010; Chen et al., 2016a). In addition to strong intracellular incompatibilities, extracellular factors can enhance the deleterious effect of some hybrid incompatibilities.

Environmental selection, even in absence of geographic barriers, is a known impetus of reproductive isolation (Hou et al., 2015, 2016; Landguth et al., 2015). This is often observed during separation of incipient species, where weak DM incompatibilities begin to emerge. In fact, mitochondrial-nuclear interactions in *Saccharomyces cerevisiae* populations can explain a significant proportion of the phenotypic variances under diverse environmental conditions. The allelic interactions between mitochondrial and nuclear genomes may be co-adapted to specific ecological niches that the yeasts occupy. Disturbing such naturally occurring interactions leads to breakdown of within-environment mitochondrial-nuclear epistasis, which provides fitness advantages in certain environments (Paliwal et al., 2014). Like nuclear genetic variations, mitochondrial genetic variations are also a source of

adaptive potential. In isogenic strains containing recombinant mtDNAs, multiple loci interact epistatically and are specific to some environmental conditions. Interruption of co-adapted mitochondria-mitochondria interactions causes fitness defects in these environmental conditions (Wolters et al., 2015, 2018). Some mitochondrial-nuclear incompatibilities in yeast become evident only under stress when the hybrids of obligate fermentative yeast are forced to respire in non-fermentative carbon sources (Hou et al., 2015, 2016). These weak incompatibilities observed between *S. cerevisiae* populations may eventually lead to strong incompatibilities similar to those existing between close relatives of this budding yeast (Chou et al., 2010; Chou and Leu, 2010; Hou et al., 2016; Jhuang et al., 2017). Hybrid necrosis in plants can also be condition specific. For example, while in some plant hybrids weakness becomes evident at high temperatures (Bomblies et al., 2007; Alcazar et al., 2009; Chen et al., 2013, 2014), other plants usually demonstrate hybrid breakdown at low temperatures (Traw and Bergelson, 2010; Hua, 2013).

In all of the above examples, there is a strong motivation to identify the specific molecular mechanisms that can give rise to hybrid incompatibilities leading to speciation events. The most-simplified DM model assumes that genetic incompatibilities are due to pairwise genetic interactions that contribute additively to hybrid breakdown between diverging lineages (Turelli and Orr, 2000; Welch, 2004). In fact, almost all of the well-studied DM incompatibilities happen to be two locus incompatibilities. Several theoretical models based on the holey adaptive landscape model and Fisher’s geometric theorem has shown that multi-locus DM incompatibilities can exist. These models predict the fitness effects of mutations on a population based on the probability that allelic interactions are incompatible, and by estimating the effect of mutations on multiple traits (Fraisie et al., 2016). Experimental crosses have frequently shown that reduction in fitness can be due to impaired interactions at more than two locus (complex epistasis) supporting the theoretical predictions (Moyle and Nakazato, 2009; Kao et al., 2010; Burkart-Waco et al., 2012; Corbett-Detig et al., 2013). However, very few examples of complex epistasis have been dissected at the genetic level (Moyle and Nakazato, 2009; Tang and Presgraves, 2009; Kao et al., 2010; Corbett-Detig et al., 2013; Phadnis et al., 2015).

Another theoretical model explaining the advent of complex epistasis is the “snowball” effect. As populations diverge the number of incompatibilities are expected to increase faster than linearly, “snowballing” the populations toward distinct species (Orr and Turelli, 2001; Turelli and Moyle, 2007; Presgraves, 2010b). Testing the “snowball” effect requires information on genetic incompatibilities between species at different divergence times. Since this information is hard to get, evidence for the “snowball” theory has been scarce. However, genetic mapping data in *Drosophila* (Matute et al., 2010) and *Solanium* species (Moyle and Nakazato, 2009) has demonstrated that accumulation of weak DM incompatibilities can “snowball” and strengthen the genetic barrier between species. The original “snowball” model (Orr, 1995) was developed for diverging populations, considering simple DM incompatibilities, and did not account for loss of DM incompatibilities when these

populations diverged. Later extensions of this “snowball” model (Kalirad and Azevedo, 2017) has shown that DM incompatibilities do not necessarily function independently of each other and several DM incompatibilities can arise and then disappear, while the populations are diverging, more so in complex incompatibilities (more than two loci) that are derived from the changes subsequent to the initial DM incompatibilities.

Factors that modulate the effect of two or more incompatible locus are allele frequency and genetic background. While changes in allele frequencies at incompatible locus can affect its relationship with its partner locus, a change in genetic background can alter the strength and magnitude of the incompatible locus (Wade, 2002; Johnson, 2010). Thus, individually some components of complex epistasis are either weak or they demonstrate incomplete penetrance (Lopez-Fernandez and Bolnick, 2007), but are synergistic, and demonstrate severe fitness defects and cause hybrid breakdown, when all the members of complex epistasis are present (Wu and Palopoli, 1994; Presgraves, 2010a; Lindtke and Buerkle, 2015; Schumer et al., 2015). As discussed earlier, these weak incompatibilities formed at the early stages of speciation may initially reduce gene flow between populations and “snowball” to become strong incompatibilities. Thus, deciphering complex epistasis can yield insights into the process of divergence at the early stages of speciation and are important for determining the rate and patterns of evolution during speciation (Kondrashov, 2003; Welch, 2004; Maheshwari and Barbash, 2011).

PROTEIN COMPLEX MICROENVIRONMENTS PROVIDE A MOLECULAR BASIS FOR MULTI-LOCUS INCOMPATIBILITIES

Proteins, the building blocks of the cell, execute many cellular functions through protein-protein interactions (PPIs). A large fraction of PPIs in eukaryotic proteomes culminate as heteromeric-protein complexes, and are responsible for diverse biochemical activities essential to cellular homeostasis, growth, and proliferation. For example, over 60% (about 3,600 proteins) of the *S. cerevisiae* proteome (Pu et al., 2009; Benschop et al., 2010; Costanzo et al., 2016), over 7,700 proteins in humans (Drew et al., 2017a), and over 2,700 proteins in fly (Guruharsha et al., 2011), are identified as subunits of protein complexes. Amino acid residues can evolve at different rates within a protein, with compensating mutations co-evolving at protein interaction sites (Zhang and Gu, 1998; Neverov et al., 2020; **Figures 1A,B**). The residues at intra-protein contact sites also co-evolve with entangled substitutions. A mildly destabilizing mutation at a protein site may lead to a compensatory mutation at its contact site that is fixed more easily and re-establishes the protein's function (**Figures 1C,D**; Weigt et al., 2009; Burger and van Nimwegen, 2010; Morcos et al., 2011; Rizzato et al., 2019). This leads to variability in accumulation of substitution types (heterotachy and heteropercilly), becoming more evident as the species diverge (**Figures 1E,F**; Lopez et al., 2002; Roure and Philippe, 2011). Such micro-evolution within a protein can lead to co-evolution of partner proteins both in

pairwise protein interactions as well as protein complexes. With individual subunits within a protein complex often sharing similar evolutionary patterns, alterations in partner proteins can decrease stability of protein complexes, create the loss of interactions, and even lead to failures in their assembly (Harrison and Burton, 2006; Juan et al., 2008). Protein-protein interactions are one of the major determinants of protein evolutionary rates and in general evolutionary rates are negatively correlated with the number of interactions (Fraser et al., 2003; Fraser and Hirsh, 2004; Fraser, 2005). Also, physically interacting proteins tend to co-evolve, and are precisely co-expressed to maintain the proper stoichiometry of interacting partners (Fraser et al., 2004). However, there are several known protein complexes with complex subunits (within conserved complexes) that still show high evolutionary rates deviating from the general trend. An example of protein complexes with such contrasting patterns of evolution would be the nuclear pore complex and RNA polymerase II (Leducq et al., 2012). Even though the underlying driving force is unclear, these proteins are the candidates that may cause incompatibilities.

Co-evolving protein-protein interactions depend on the phenotypic traits, which are species specific and maintained by reciprocal selection (Brodie and Ridenhour, 2003; Klink and Bazykin, 2017). These principals of protein co-evolution have been exploited to predict protein-protein interactomes at a genome scale and also in protein structure predictions (Marks et al., 2012; Hopf et al., 2014; Anishchenko et al., 2017; Cong et al., 2019). Thus, the divergence of orthologs proteins between species or populations can manifest as incompatibilities in hybrids, when co-evolving partner proteins gets shuffled/swapped from the two parents in the hybrids (Clark et al., 2009; Tang and Presgraves, 2009; Zamir et al., 2012; Ochoa and Pazos, 2014). Furthermore, proteins often fold into functional conformation only after interacting with their partners (Dunker et al., 2008). Loss or failure in interactions can lead to protein misfolding and loss of its stability. Interacting partner proteins thus tend to co-evolve with each other within a species (Neverov et al., 2020) and are also often co-expressed with one another, to maintain proper stoichiometry among interacting components (Ge et al., 2001; Fraser et al., 2004). This is more evident in protein complexes, as multiple proteins are required to interact, and the functional folded conformation of protein complexes may depend on some of these interactions.

MIS-ASSEMBLY OF PROTEIN COMPLEXES CAN BE A SOURCE OF MULTI-LOCUS INCOMPATIBILITIES

Most protein complexes are assembled co-translationally, unidirectionally (Gloge et al., 2014; Shiber et al., 2018; Schwarz and Beck, 2019) and in a specific order (Marsh et al., 2013). Protein complex assembly is a multistep process and must proceed *via* the most energetically optimum path of bringing together proteins to form intermediate subcomplexes, which is further extended to form complete protein complexes (**Figure 2A**). By analogy to Levinthal's paradox of protein

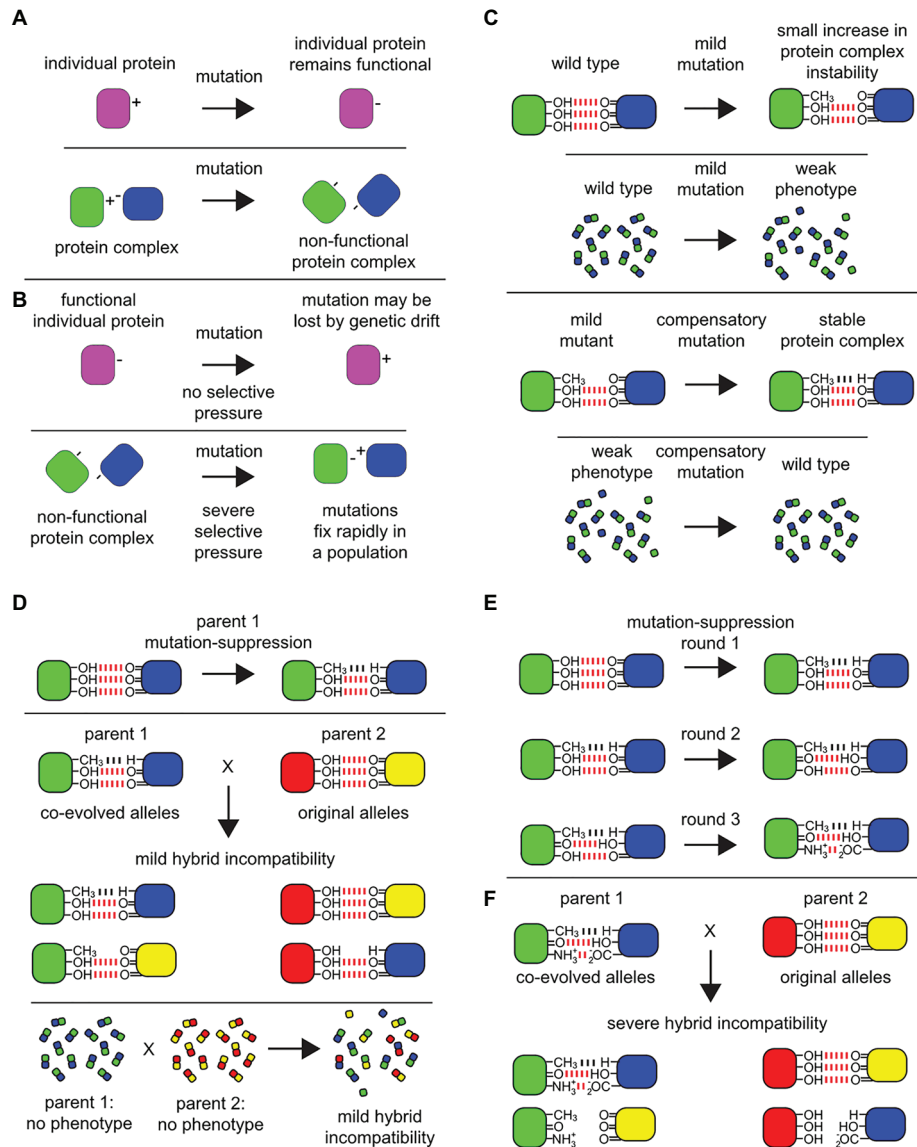
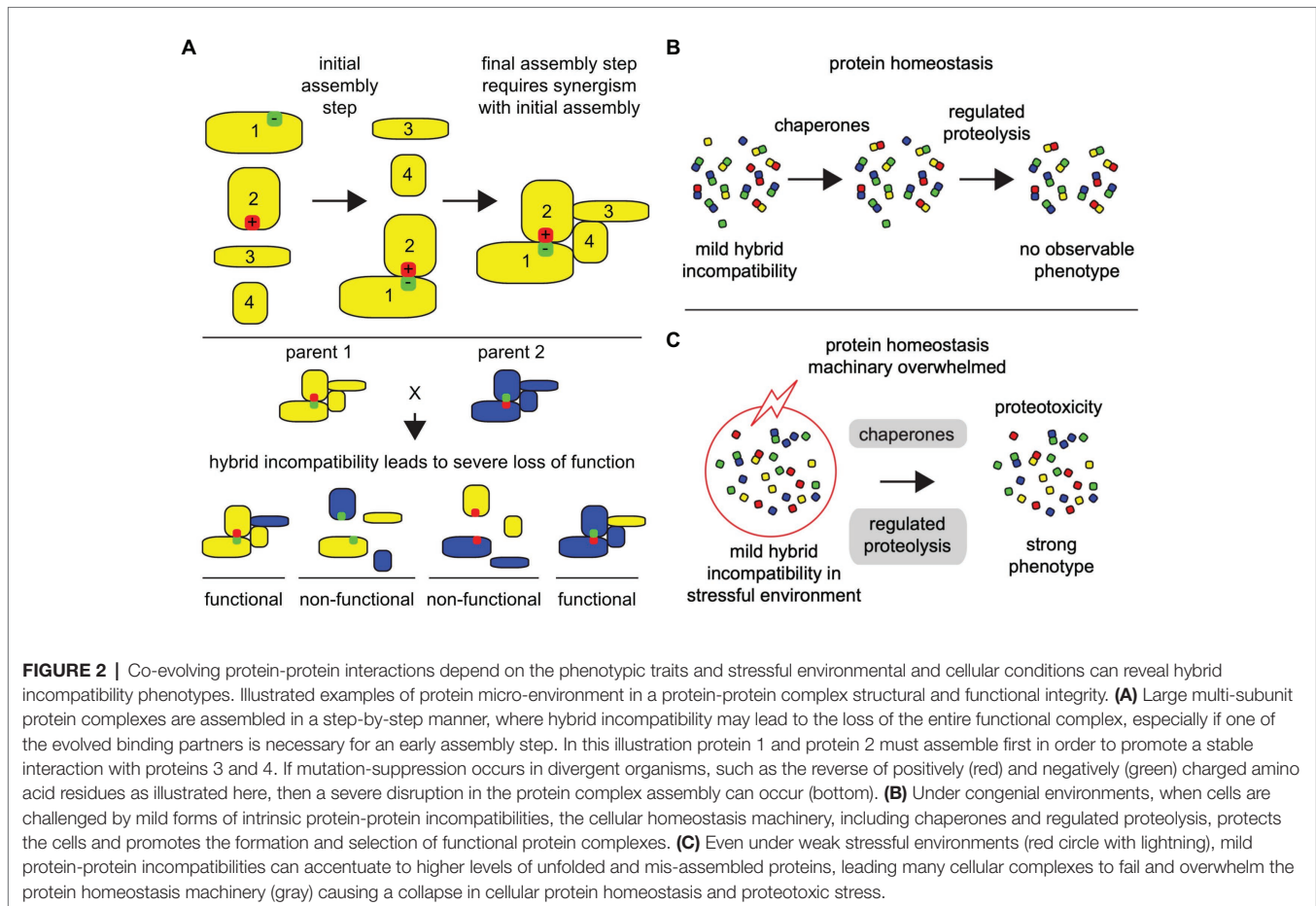


FIGURE 1 | Protein complex micro-environments provide a molecular basis for multi-locus incompatibilities. Among biomolecular physical interactions within cells that give rise to hybrid incompatibilities, protein-protein interactions (PPIs) are a rich target for experimental investigation. **(A)** An example of a mutation on a monomeric protein versus a similar mutation at the interface of a PPI. A monomeric protein (magenta) undergoes a mutation at a surface residue from a positively charged amino acid residue to a negatively charged one which displays no phenotype and is fully functional. By contrast, in a protein-protein complex (green-blue) a mutation in one of the subunits (green) at a surface amino acid residue within the interaction domain from a positively charged residue to a negatively charged one could cause a loss of function, which may display a phenotype subject to selective pressure. **(B)** The mutation in the monomeric protein may be lost over time due to genetic drift. By contrast, among the subunits of the non-functional protein-protein complex, under severe selective pressure, a compensatory mutation may occur in the binding partner protein (blue) to allow for the formation of a function complex where both mutations rapidly fix within the population. However, note that the charge residues on the surface binding sites have been reversed between the subunits. **(C)** An example of a mild mutation within a protein-protein interaction site. An interaction between subunits in a multi-protein complex are stabilized by non-covalent bonds, where in this example three hydrogen bonds (red) create the binding energy to stabilize the complex. The fictional mutation illustrated here leads to the loss of one hydrogen bond in the protein-protein interaction domain of the complex. Within the cell, a mild mutation may lead to a small decrease in the average stability of the protein complex, illustrated in this diagram as a failure of only a few protein complexes to maintain their structure, yielding a very mild phenotype. Under selective pressure over time, compensatory mutations can occur in the binding partner to suppress the mild defect and fix within the population (bottom), illustrated here as a Van der Waals potential (black). **(D)** Co-evolved binding partner alleles can be uncovered as incipient mild hybrid incompatibility, which may or may not give rise to a detectable phenotype in hybrids. Even if there is no detectable phenotype, at the quantitative level among the population of individual proteins and protein complexes within the hybrid cells, this incipient incompatibility can lead to failures in the proper assembly of the protein complex and/or lead to a decrease in the stability of the protein complex (bottom). **(E)** After multiple rounds of mutations that display a mild phenotype, followed by compensatory suppressor mutations in the binding partner, multiple changes in the micro-environment in the protein-protein interaction domains can accumulate and fix in the population. **(F)** The fixation of these mutations in the protein-protein interaction domains can be revealed later as a source of hybrid incompatibility.



folding (Levinthal, 1969), just as proteins fold *via* limited number of energetically favorable folding pathways, protein complexes should be expected to assemble following a strict order and known to be under evolutionary selection (Marsh et al., 2013). Deviation from the energetically optimum order can lead to mis-assembly of proteins with severe biological consequences (Dobson, 2003; Ellis, 2007). The assembly and maturation of protein complexes are coordinated by molecular chaperones. Generally, protein homeostasis machinery and related chaperones closely monitors and prevents the formation of misfolded protein aggregates (Asher et al., 2006; Hartl et al., 2011), protein trafficking, and enzyme activity regulation (Figure 2B; Mashaghi et al., 2016). The chaperone proteins require the help of co-chaperones and assisting proteins in assembling the protein complexes (Marsh and Teichmann, 2015). One of the ubiquitous family of chaperones important for protein folding and protein complex assembly are the heat shock proteins (HSPs). Chaperoning and assembly of protein complexes by HSPs has been studied in considerable detail (McClellan et al., 2007; Li et al., 2012; Makhnevych and Houry, 2012; Gopinath et al., 2014; Marsh and Teichmann, 2015). First, Hsp70, Hsp40, and client proteins form an early complex, and this is transferred to Hsp90 with the help of the adaptor protein Hop/Sti1 for correcting late stage misfolding, and final assembly.

In hybrids, which are the outcome of the mating of parents from related species, the protein complexes are formed from an admixture of subunits from the two parents. Under congenial environmental conditions, the homeostasis machinery in hybrids derived from closely related species or populations works toward maintaining the protein-protein interactions and protein complexes, at least in a partial functional state (Figure 2B; Leducq et al., 2012; Piatkowska et al., 2013; Zhong et al., 2016). However, when these closely related hybrids are exposed to even mildly harsh external environments, the protein homeostasis machinery is overburdened and leads to breakdown of interaction between proteins (Figure 2C). This has been evidenced in several studies (Leducq et al., 2012; Zill et al., 2012; Piatkowska et al., 2013). Mis-assembly of protein complexes due to failures in protein-protein interactions can impair the protein homeostasis machinery invoking a proteotoxic stress response, and can cause severe growth defects (Gidalevitz et al., 2011; Arslan et al., 2012; Oromendia and Amon, 2014; Radwan et al., 2017). This is also true for essential proteins. Several essential proteins are subunits of protein complexes and are incompatible as chimeric assemblies from closely related species (Lai et al., 2018). Proteotoxic response in hybrids could be due either or all of the above-described reasons, i.e., loss of interactions among co-evolving proteins, reduced protein expression, mis-folding of proteins to non-functional configurations.

Protein complexes involve multiple interactions and can result in different levels of functional defects depending on how many interactions are compromised. Thus, phenotyping chimeric protein complexes in parental hybrid cellular environments can enhance our insight into the molecular basis of emergence and fixation of multi-locus incompatibilities. This requires determining and analysis of large-scale proteomics data in parental species as well as their hybrids.

In addition, mis-regulation of protein complexes in hybrids could also be due to transcriptional mis-regulation or mRNA instability (**Supplementary Figure 1**). Divergence between interacting regulatory elements or regulatory divergence is another common route through which DM incompatibilities can arise (Mack and Nachman, 2017). mRNA abundance is regulated by the binding of *trans*-factors (mainly Transcription Factors) to *cis*-regulatory elements, which are short stretches of non-coding DNA, thus mutations on either of them can affect the mRNA abundance of target genes. While *trans*-factors are known to be under higher selection constraint than their *cis*-counterparts, they can evolve faster than other classes protein coding genes (Castillo-Davis et al., 2004). Transcriptional regulation diverges quickly between closely related species that often leads to mis-regulated gene expression in hybrids (Landry et al., 2005; Tirosch et al., 2009; Tirosch and Barkai, 2011; Swain Lenz et al., 2014; Mack and Nachman, 2017; McGirr and Martin, 2019). Although, inter-species transcriptomic analysis has shown that changes in transcript levels are frequently deleterious (Gilad et al., 2006), gene regulatory networks are not necessarily conserved between species (True and Haag, 2001). The evolution of gene expression can be explained under a “house of cards” model of stabilizing selection, where mutations affecting mRNA abundance can lead to a deluge of changes between co-evolved *cis*-elements and *trans*-elements in an evolutionary network. Such evolutionary cascade has been observed in yeast, worms and flies (Hodgins-Davis et al., 2015). Further details on the theoretical and empirical considerations of gene regulation and their implications on speciation can be found in a recent review by Mack and Nachman (2017).

Promoter-mediated coupling of transcription to mRNA degradation is known to be diverged between closely related species (Dori-Bachash et al., 2012). Altered rates of mRNA degradation can impact dosage of available transcripts for translation affecting protein expression and abundance. The translational efficiency of proteins also changes with alterations in intrinsic and extrinsic environmental conditions. One of the mechanisms by which the translational machinery adapts to such changes is by modulating their tRNA usage (Meiklejohn et al., 2013; Yona et al., 2013; Bloom-Ackermann et al., 2014). The translational efficiency is dependent on cellular concentration of tRNA molecules and the efficiencies of each codon-anticodon pairing (Dos Reis et al., 2004). There is an apparent connection between tRNA availability and protein folding. Since, tRNA usage is fine-tuned within each species, changes in tRNA pool in interspecies hybrids can affect their translational efficiency and protein folding (Meiklejohn et al., 2013; Yona et al., 2013; Bloom-Ackermann et al., 2014). When one or more subunits of a protein complex is mis-regulated due to anyone of the

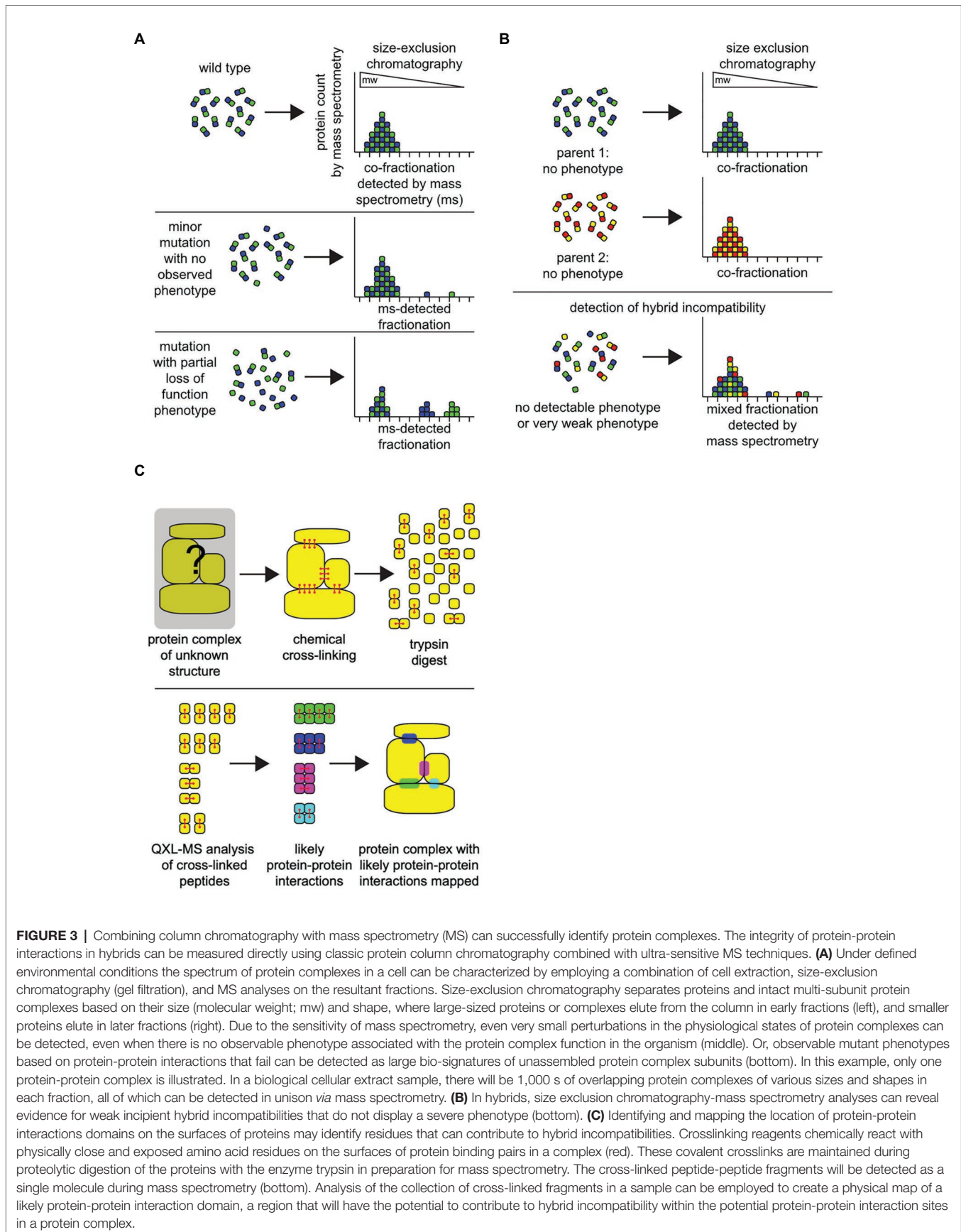
above reasons or their mRNA is degraded, dosage of available transcripts for it is altered for these subunits. This can interfere with the tight stoichiometry with which protein complex subunits are produced, leading to protein mis-assembly. Furthermore, DM incompatibilities due to regulatory divergence can also lead to breakdown in protein complexes. Protein complexes can be compared to a jigsaw puzzle, with each subunit forming an element of the puzzle. Misfolded or missing complex subunits might not be able to complete this puzzle leading to incomplete protein complex assembly.

PROTEOME-SCALE ANALYSES OF PROTEIN COMPLEXES

Protein complexes are usually identified by characterization of PPI. Some of the commonly used experimental strategies to detect PPIs, include the yeast two-hybrid system (Y2H; Paiano et al., 2019), protein-fragment complementation assay (PCA; Tarassov et al., 2008; Michnick et al., 2010), fluorescence resonance energy transfer (FRET; El Khamlichi et al., 2019), and affinity purification plus mass spectrometry (AP-MS; Gavin et al., 2006; Krogan et al., 2006; Volkel et al., 2010; Babu et al., 2012; Adelmant et al., 2019). These methods help us collect a vast set of protein-protein interaction data from different organisms. However, most of these approaches are limited by the availability of high-quality antibodies or sequence-verified cDNA clones suitable for targeted protein complex enrichment, labor-intensive and not suitable for the global complex dynamic studies in which many time points and conditions are involved.

Recent advances in mass spectrometry techniques allow us to develop an effective approach to quantitatively measure PPIs. To gain a global view of the complex behavior, transient and higher-order associations of the proteome under a state mimicking the native cellular condition needs to be captured. Some of the recent studies have successfully shown that cellular fractions could serve as a proxy for the cellular environment and retain basic cellular organization especially for the non-membrane proteins (Kristensen et al., 2012; Kastritis et al., 2017; **Figure 3A**). Such approaches have also proven to be effective in non-model organisms. A recent study combined size exclusion chromatography and ion-exchange chromatography with tandem mass spectrometry to successfully identify protein complexes in five metazoan species *Caenorhabditis elegans* (worm), *Drosophila melanogaster* (fly), *Mus musculus* (mouse), *Strongylocentrotus purpuratus* (sea urchin), and human. The identified protein complexes were compared with the data derived using a similar co-fractionation strategy in *Xenopus laevis* (frog), *Nematostella vectensis* (sea anemone), *Dictyostelium discoideum* (amoeba), and *S. cerevisiae* (yeast) to gain insight into biochemical evolution of protein interactions and conservation of protein complexes in species diverged by over billion years (Wan et al., 2015).

Although the co-fractionation strategy by itself has been found to be effective, it still lacks sensitivity for in depth analysis of protein complex dynamics. The sensitivity in



identification of proteins and its complexes can be improved significantly with stable isotope labeling by amino acids in cell culture (SILAC; Ong et al., 2002). By incorporating the SILAC protocol, samples from different conditions or time points can also be accurately quantified and compared. Since SEC separates the protein complexes based on their size, the subunits of an assembled complex should be observed in the same fractions (Kastritis et al., 2017). If the same complex subunits appear in different fractions under different condition, it indicates that the complex structure has been altered or destabilized (**Figure 3A**). There are only a handful of studies that have attempted to determine protein complexes in hybrids (**Figure 3B**; Leducq et al., 2012; Piatkowska et al., 2013; Bontinck et al., 2018). Most of these studies have focused on specific protein complexes in related species. An experimental limit of the SILAC approach is that cells or organisms need to be cultured in the isotope-containing medium before the samples are prepared, thus non-applicable to the systems that cannot be cultured in the lab. In recent years, the efficiency of isobaric labeling methods, such as tandem mass tags (TMT), have been greatly improved (Thompson et al., 2003; Wang et al., 2020).

Although the *in vitro* labeling process inevitably introduces more variation into the samples, it provides an alternative approach for accurate measurements of protein complex dynamics in unculturable systems. The co-fractionation approach to detect the potential basis of complex hybrid incompatibility is also limited by the need for exceptional rigor and caution during sample preparations, as the signature for incompatibility is the disruption of protein-protein interactions as detected by SEC/MS, which can also easily result from mis-handling of samples. In general, very small differences in sample preparations and sample handling can lead to disparate results, where it remains unclear which sample(s) contain the most accurate representation of the spectrum of protein complexes *in vivo* (Ho et al., 2002). One set of fundamental unavoidable physical variables contributing to these uncertainties is simply that the act of cell lysis always dilutes the concentration of the proteins into a new solution environment that does not precisely match the native cytosol, where this disruption in protein complex homeostasis is also amplified by the inherent errors in human liquid handling through time, and on milliliter or microliter scales. Automated nano-volume high-precision rapid cell extraction may help to suppress variation in sample preparation. Toward this end, one promising technical development is nano-scale single-cell mass spectrometry technologies, where, in the future, it may be possible to couple these nano-liquid-handling approaches with a size-fractionation step in order to map out the distributions of protein-protein interactions with high-accuracy and reproducibility (Williams et al., 2020). Alternatively, before cell extraction, or immediately upon cell extraction, covalent chemical cross-linking can be employed to both stabilize protein-protein interactions and to potentially map interaction domains, but has the drawback that non-specific cross-linking can occur leading to false-positive results.

In general, co-fractionation mass spectrometry (CF-MS) can only be used to infer protein complexes when the information of protein functions or interactions is available. On the other

hand, quantitative cross-linking mass spectrometry (QCLMS/QXL-MS) can be used to directly detect protein interaction sites and binding partners (**Figure 3C**). Reactive groups in proteins and protein complexes are cross-linked and subjected to mass spectrometry. The mass spectrometric signals of cross-linked peptides derived from different conformations can then be distinguished. The location of the cross-links inflicts a distance constraint on the respective side chains orientation and position. Thus, it can be used to draw conclusions on the three-dimensional structure of the protein or topology of a protein complex (Sinz, 2006; Chen et al., 2016b; Kastritis et al., 2017; Chen and Rappsilber, 2019). The recent advances of QXL-MS allows us to construct the global protein complex list in the organisms for which information of protein functions or interactions are limited (Chavez et al., 2019). Moreover, it can be used to measure the change of protein-protein interaction systematically or detect chimeric protein complexes in hybrids.

Current high-throughput proteomics platforms come with several software tools for pre-screening, identification, assembly, and quantification of detected peptides. Such built-in tools are usually tailored for optimization of the data generated by mass spectrometers depending on their fragmentation processes and collision energies and generally do not work across platforms. MASCOT, OMSSA, SEQUEST, X!Tandem, and TOPPAS are some of the popularly used platforms for analyzing tandem mass spectrometry data. The popular tools that are used for analysis of peptides generated from SILAC or QCLMS/QXL-MS followed by LC-MS/MS are Census, MaxQuant, MsQuant, MASCOT Distille, and COFRADIC (Kohlbacher et al., 2007; Nesvizhskii et al., 2007; Cox et al., 2009; Deutsch et al., 2010; Nahnsen et al., 2011; Doerr, 2012; Kristensen et al., 2012; Yates et al., 2012). One of the major draw backs of proteomics data is that the signal to noise ratio is extremely small. The tools developed till-date have been developed to enhance the signal and reduce the noise in the data. Filtering noise and identifying species specific proteins in proteomics data from hybrid species, especially from closely related species, is challenge, leading higher false positive rates in the predictions. Fractionation based on size and isotope labeling of amino acids followed by tandem mass spectrometry has improved the resolution of separation of peptides, but there is still significant loss of data in hybrids due to lack of specificity in peptides between the two species forming the hybrid. However, in the recent years, considerable effort has been devoted to improving the experimental and analysis techniques of mass spectrometry and the limitation of proteomics can be expected to be alleviated in the near future. This is corroborated by several studies in the last few years, where global analysis of protein complexes has been done in a wide range of higher order biological systems including plants (Aryal et al., 2014), pluripotent stem cells, and cancer cells (Sudhir and Chen, 2016). The technology has also advanced to handle large scale quantitative proteomics; proteins in 375 cancer cell lines were quantitatively profiled recently using mass spectrometry (Nusinow et al., 2020). Developing experimenting techniques and the analysis software that is specifically applicable

to hybrid analyses is an area of research with definitive scope for innovation and development.

IN SILICO PREDICTION OF PROTEIN COMPLEXES

Almost all the available methods for predicting protein complexes rely on protein-protein interactions. Experimental data are the primary source of these protein-protein interactions. Existing prediction methods can be classified into network-based approaches and functional information-based approaches. Although, methods from the two categories are used independently to predict protein complexes, they are also frequently used in combination (Price et al., 2013; Srihari and Leong, 2013; Zahiri et al., 2020). The network-based approach classifies protein-protein interactions as protein complexes based on the density of protein interactions, and topology of network structure. The functional information methods, also known as biological context-based methods, supplements the protein interaction network with information from functional annotation such as gene expression, gene ontology, and protein domain architecture. Supplementing annotation is known to improve the accuracy of predictions more than those based only on pure protein network approaches (Price et al., 2013; Srihari and Leong, 2013; Zahiri et al., 2020).

Accurate predictions of network topologies require accurate affinity scoring schemes. The most accurate whole organism complexome is derived from high-throughput TAP-MS studies, based on the affinity scoring schemes developed specific to their experimental design and output (Gavin et al., 2006; Krogan et al., 2006; Collins et al., 2007; Hart et al., 2007). The affinity scores can be used to determine the confidence and reliability of observed protein-protein interactions. Availability of accurate affinity scoring schemes from TAP-MS data is currently limited to a few standard model organisms such as yeast *S. cerevisiae*. In higher order organisms and non-model species direct protein-protein contacts can be derived using the from the covariation pattern of the protein abundances from CF-MS data. The three-dimensional arrangement of protein contacts leading to protein complexes can be determined by performing clustering analysis of correlation matrices derived from CF-MS profiles using machine learning algorithms. The CF-MS system has been successful in identifying complexomes of *Caenorhabditis elegans*, human cell lines, and 13 plant species of agricultural and scientific importance (Drew et al., 2017b; Hu et al., 2019; McWhite et al., 2020). The CF-MS correlation profiles can be considered analogs to affinity scores from TAP-MS (henceforth referred to as affinity scores for simplicity). Markov Clustering algorithms (MCL) have proven to be reasonably successful as a method for incorporating information from multiple resources and developing weighted protein interaction networks (Srihari and Leong, 2012). MCL is an unsupervised cluster algorithm, which simulates a series of random walks and iteratively computes the probability that the cluster of protein interactions is dense or sparse during each visit (Enright et al., 2002; Van Dongen and Abreu-Goodger, 2012). In unsupervised clustering algorithms, the data is not labeled and the model is allowed

to work on its own to group data inherently into clusters. When the probability of cluster is dense, the MCL will not leave the cluster. Based on the thickness and spread of the random walks to a cluster, MCL identifies protein complexes.

The basic MCL works well for small clusters but is not successful for clusters with more than 10 partners. This bias to prediction of several small clusters (fragmentation bias) has been resolved to a large extent by advance variants of MCL such as Multi-level regularized MCL (Satuluri and Parthasarathy, 2009; Satuluri et al., 2010) and Parallel Shotgun Coarsened MCL (Lim et al., 2019). These methods start with initial clusters using MCL and then build bio-networks of refined clusters by incorporating core-attachment structures to generate complexes. Detecting core-attachment happens in two stages (Wang et al., 2019). First, a subgraph with maximum clique, i.e., largest complete subgraph with all its vertices connected to each other, are identified as a protein core. Next, protein attachment structures are determined by selecting proteins that interact with more than half of the neighboring protein cores.

Despite these advances, the false positive rate in protein complex prediction is high. This is mainly due to high variability in complex core sizes and the sparsity of protein-protein interaction networks. Also, the protein complex compositions can vary between experiments for the same strains of the same species. These differences can be due to the experimental procedures, dissimilar coverage in data from different experiments and from the application of pre-screening software. Thus, it can lead to an ambiguity on the correct composition of protein complexes. To overcome these limitations, studies have tried to include functional information of interacting proteins. This information includes protein domain architecture, protein expression, gene expression, and gene ontology. Recent methods have indicated that including evolutionary information and genetic algorithms can improve the accuracy in protein complex predictions (Zahiri et al., 2020). Bayesian networks and other probabilistic classifiers developed using the parameters derived from MCL and its variants along with the functional parameters have also been found to be reliable (Gavin et al., 2006; Krogan et al., 2006; Collins et al., 2007; Hart et al., 2007; Wang et al., 2009a). These classifiers are also useful in determining protein complexes from protein contacts derived from CF-MS data. Although, evolutionary algorithms and Bayesian classifiers are more accurate, they are computationally expensive. While, considerable work has been devoted to optimizing Bayesian classifiers, evolutionary algorithms although superior in their predictions, are yet to be optimized for efficient application on large networks.

Furthermore, considerable effort has been devoted to develop algorithms for compiling core consensus protein complexes from diverse data resources. For example, Benschop et al. (2010) developed a forward-backward module detection algorithm that predicts consensus of protein complexes from AP-MS data in *S. cerevisiae*. This algorithm searches across the diagonal on the clustered protein interactome matrix for strongly interacting proteins, by traversing first from right to left and then from top to bottom. This is the forward phase of the algorithm, in the backward phase the direction of traversal is reversed, and detects and accounts for non-interacting proteins.

Similar methods have also been developed for predicting consensus soluble protein complexes isolated from human HeLa S2 and HEK93 cells (Havugimana et al., 2012) and *Arabidopsis thaliana* (Gorka et al., 2019).

Bioinformatics methods have also played an important role in advancing the field of proteomics into more complex areas. For example, the determination of composition of membrane protein complexes, have been elusive as upon employing detergents, which is necessary to elute membrane proteins, protein complexes fall apart. Recent studies have determined the protein complexes based on the protein interaction maps in *Escherichia coli*, *S. cerevisiae*, and human mitochondria (Babu et al., 2012; Maly et al., 2017; Babu et al., 2018). They have also helped in determining the biochemical evolution of protein complexes diverged by over a billion years (Wan et al., 2015). Together, these demonstrate the importance in advancement of bioinformatics tools and algorithms in the field of proteomics and predicting protein complexes. However, most of the currently available bioinformatics tools are designed for pure species and mainly tuned to work efficiently on data from a handful of model organisms. The underlying reason is that the availability of proteomics and protein-protein interaction data is also limited to this subset. Thus, there is a huge void in both data and techniques to deal with non-model organisms, and, more importantly, hybrid organisms.

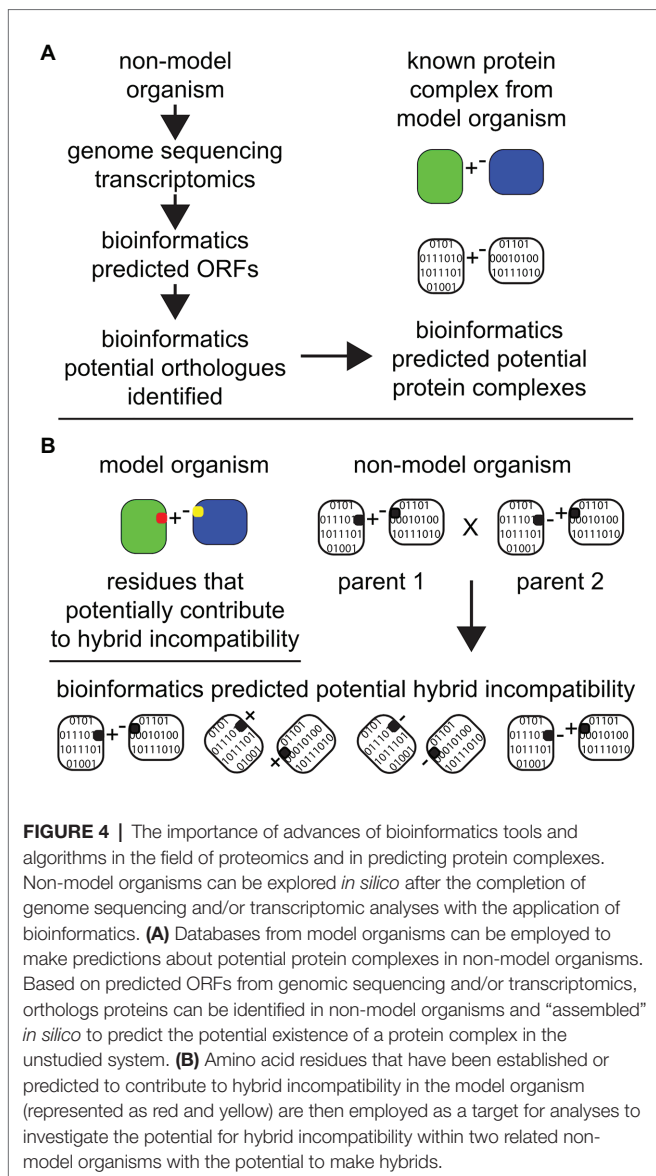
GENOMES AND TRANSCRIPTOMES CAN ASSIST PREDICTING GENETIC INCOMPATIBILITIES IN NON-MODEL SPECIES

Next Generation Sequencing (NGS) technologies have the potential to work with any species, determine the natural variation at a genome-wide level and at unprecedented resolution, and also can provide us to with a comprehensive picture of regulatory variation. Over the past decade, such genomic technologies have been used to explore the extent of natural variability at the molecular level and the evolutionary forces that shape this variation (Gilad et al., 2009; Wolf et al., 2010). In several cases, patterns of variation at either structural or regulatory levels have helped in explaining physiological and morphological phenotypes (Hoekstra and Coyne, 2007; Carroll, 2008). NGS has also been able to discover genetic incompatibilities in *taxa* at their incipient stage of speciation and introgressed species (Barreto et al., 2011) as well as non-model species (Gagnaire et al., 2012).

Resequencing and *de novo* assembly are two frequently used modes of genome assembly. Resequencing refers to sequenced reads performed when a reference genomes are available and *de novo* assembly is used when the reference genome is of poor quality or when no reference genome is available. *De novo* assembly is more computationally intensive compared to resequencing but has wider applications. The quality of a genome assembly provides a measure of the degree to which the sequence has been correctly assembled and the sequences are reliable, and thus of great importance. Assembly quality can be assessed using different statistics, which offer a measure of genome

completeness and contiguity (Yandell and Ence, 2012; Lachance and Tishkoff, 2013; Simao et al., 2015; Liao, 2019). Excellent reviews are available on *de novo* genome assembly (Liao, 2019) and use of genome sequencing in non-model organisms (Ellegren, 2014). The recently developed third generation sequencing technologies can produce extra-long reads with a median length of 10–20 kb and sometimes longer than 50 kbp. Such long reads can facilitate better quality of *de novo* assemblies especially in non-model organisms (Giani et al., 2020). Similarly, in RNA-Seq reads can be assembled *de novo*, as well as mapped to a reference genome or transcriptome and used to quantify gene expression changes between the organisms of interest (Wang et al., 2009b; Oshlack et al., 2010). In addition, transcriptomic data from RNA-Seq can also be used for population genetic analyses, test for selection or isolate SNPs/microsatellite markers for further population genetic or genetic mapping studies (De Wit et al., 2012; Sun et al., 2012).

Current genetic and molecular biology strategies are confined to two or three locus incompatibilities due to practical and technical restraints. Determining multi-locus incompatibilities is difficult due to the high level of epistasis in hybrids (Kao et al., 2010; Li et al., 2013). Since phenotypic traits due to genetic interactions (epistasis) can appear in ratios deviating from those expected with independent assortment, it becomes more difficult to detect multi-locus incompatibilities. Predicting and validating multi-locus incompatibilities requires considerably larger sample sizes and is a major hurdle in traditional approaches. Transcriptomic and genomic information can help to generate possible protein interaction networks that can further be used to alleviate the complexity of epistasis (Figure 4; Angeles-Albores et al., 2018). There are a few reviews elaborating the importance of integrative bioinformatic approaches to incorporate genomic, transcriptomic, and protein-protein interaction data along with their merits and demerits (Nesvizhskii, 2014; Wang et al., 2014; Wang and Zhang, 2014; Ruggles et al., 2017). Transcriptomic data has been previously used for functional annotation of identified proteins and determining protein interactions leading to protein complexes. This is based on a transcriptome-interactome correlation mapping strategy, i.e., the expression profiles of interacting protein coding genes are correlated and this knowledge can be leveraged to predict protein-protein interactions (Figure 4; Ge et al., 2001; Jansen et al., 2002). This methodology has been extended to also predict protein complexes (Zhang et al., 2004; Ruggles et al., 2017; Will and Helms, 2019). Functionally conserved orthologs protein coding genes have similar expression patterns (Chen and Zhang, 2012; Rodriguez-Cruz et al., 2013; Das et al., 2016), and this can be used to predict protein complexes in related non-model species. Furthermore, the evolutionary rates of the interacting protein partners are known to be constrained with respect to each other (Fraser et al., 2003; Qian et al., 2011; Choi and Hannenhalli, 2013; Schoenrock et al., 2017). Together, the transcriptomic data coupled with evolutionary rate analysis derived from genomes can thus help in increasing the confidence in protein-protein interactions (Schoenrock et al., 2017; Hawe et al., 2019), and to a certain extent protein complexes especially when working with new model systems. Protein complexes,



which are constituted of multiple protein interactions, can contain fast-evolving proteins and thus have a higher chance of being incompatible in a hybrid cellular environment. Although more work needs to be devoted to harvest the benefits of this integrative approach, molecular evolutionary analysis of transcriptomic and genomic data show promise in studying ecological molecular speciation mechanisms (Stinchcombe and Hoekstra, 2008; Lee et al., 2014; Ravinet et al., 2017). This is specifically useful for systems where molecular and genetic techniques are inadequate to experimentally detect the genetic barriers of gene flow.

DISCUSSION

In the long term, the goal is to develop experimental and computational tools that can confidently identify the molecular

drivers of speciation. These changes can be at the genomic, transcriptomic, or at the proteomic levels. Since Charles Darwin, who illustrated “a tree of life” based on morphology, we have strived to further refine the branches in the “story of life on earth” with the addition of our knowledge about the biochemistries of metabolism and physiology, tissue histology and cellular structures, sub-cellular structures and organization, protein complexes, and with the recent great advances in genomic and transcriptomic sequencing. In addition to genetic incompatibilities and failures in protein interactions leading to a breakdown in protein complexes; protein-DNA and protein-RNA interactions (Lee et al., 2008; Chou et al., 2010; Henault and Landry, 2017; Jhuang et al., 2017) are also known causes of hybrid breakdown, albeit unexplored. These are other directions in which advances in experimental and computational techniques need development.

In addition to identifying the molecular bases of past speciation events, the aim is to see within genomes, proteomes, and protein complexes the signatures that may create blocks and barriers in future potential evolutionary trajectories. For example, the inner membrane of the mitochondria has one of the highest known protein:lipid ratios, which is proposed to make membrane integrity relatively fragile among biological membranes. It has been observed in yeast that almost all of the known cases of DM incompatibilities are mitochondrial-nuclear incompatibilities, and that there is a highly conserved AAA-ATPase homeostasis machinery essential for maintaining the integrity of the inner membrane (Francis and Thorsness, 2011; Lee et al., 2017). This suggests that in yeast the complexes of the inner mitochondrial membrane, such as the F1-F0 ATPase, are a “molecular branch point” in the natural history of yeast.

In the longer term, our knowledge about how past speciation events may have occurred in the “molecular story of life,” combined with a new potential ability to identify the potential barriers in possible future evolutionary trajectories, may allow us to bioengineer new species beyond these barriers and create organisms with increased fitness that Mother Nature has failed to create by chance over billions of years of evolution. One possible example is to bioengineer into yeast the vertebrate F1-F0 ATPase c-ring subunits, which only contains eight subunits compared to the 10 c-ring subunits of the yeast F1-F0 ATPase (Song et al., 2018), which means the vertebrate form is more efficient (Watt et al., 2010), and then study using phenotypic, molecular, and proteomic assays under diverse conditions to decipher when protein-protein interactions between the ring complex and associated proteins can increase or decrease compatibility and functionality. Laboratory evolution of hybrids between post-zygotically isolated but closely related species or cells carrying replaced orthologs from closely related species, followed by proteome analysis can help us understand how speciation barrier can be overcome and also the role of protein interactions in maintaining species integrity. For example, laboratory evolution of *S. cerevisiae* cells introgressed with CCM1-*Saccharomyces bayanus* helped discovering the general rules underlying PPR domain evolution (Jhuang et al., 2017). Laboratory experimental

evolution refinement may also allow us to create novel yeast species that may outcompete the parental strain in a way never achieved in nature during billions of years of natural selection.

AUTHOR CONTRIBUTIONS

KS and J-YL conceived the manuscript. KS, SCS, and J-YL wrote the manuscript and obtained funding. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by Ahmedabad University Startup grant to KS (AU/SUG/SAS/DBLS/19-20/01-K.BS.SWAMY_05.22), Chang Gung Memorial Hospital (CGMH), grant numbers CMRPD1J0121, CMRPD1K0121, and BMRPC59 to SCS; Academia Sinica of Taiwan, grant numbers AS-IA-105-L01 and

AS-TP-107-ML06, and the Taiwan Ministry of Science and Technology, MOST108-2321-B-001-001, to J-YL.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.609766/full#supplementary-material>

Supplementary Figure 1 | Transcriptional regulation diverges quickly between closely related species that often leads to mis-regulated gene expression in hybrids. mRNA abundance is regulated by the binding of *trans*-factors (mainly Transcription Factors) to *cis*-regulatory elements, where mutations in either of them can affect the mRNA abundance. **(A)** Ancient related species well adapted to a common ancient environment can rapidly evolve to adapt to different new environments *via* mutations that alter mRNA transcription levels and/or mRNA stability. **(B)** The ancient parents displayed hybrid compatibility, where the hybrid mRNA regulatory networks achieved the proper amount of protein production well adapted to their common ancient environment. **(C)** The rapidly evolved incipient species display hybrid incompatibility and are not well adapted to either of the current environments. The mixed regulation within the hybrids yields an average level of mRNA leading to the improper amount of protein that is not well suited for either current environment.

REFERENCES

- Adelmant, G., Garg, B. K., Tavares, M., Card, J. D., and Marto, J. A. (2019). Tandem affinity purification and mass spectrometry (TAP-MS) for the analysis of protein complexes. *Curr. Protoc. Protein Sci.* 96:e84. doi: 10.1002/cpps.84
- Alcazar, R., Garcia, A. V., Parker, J. E., and Reymond, M. (2009). Incremental steps toward incompatibility revealed by *Arabidopsis* epistatic interactions modulating salicylic acid pathway activation. *Proc. Natl. Acad. Sci. U. S. A.* 106, 334–339. doi: 10.1073/pnas.0811734106
- Angeles-Albores, D., Puckett Robinson, C., Williams, B. A., Wold, B. J., and Sternberg, P. W. (2018). Reconstructing a metazoan genetic pathway with transcriptome-wide epistasis measurements. *Proc. Natl. Acad. Sci. U. S. A.* 115, E2930–E2939. doi: 10.1073/pnas.1712387115
- Anishchenko, I., Ovchinnikov, S., Kamisetty, H., and Baker, D. (2017). Origins of coevolution between residues distant in protein 3D structures. *Proc. Natl. Acad. Sci. U. S. A.* 114, 9122–9127. doi: 10.1073/pnas.1702664114
- Arslan, M. A., Chikina, M., Csermely, P., and Soti, C. (2012). Misfolded proteins inhibit proliferation and promote stress-induced death in SV40-transformed mammalian cells. *FASEB J.* 26, 766–777. doi: 10.1096/fj.11-186197
- Aryal, U. K., Xiong, Y., McBride, Z., Kihara, D., Xie, J., Hall, M. C., et al. (2014). A proteomic strategy for global analysis of plant protein complexes. *Plant Cell* 26, 3867–3882. doi: 10.1105/tpc.114.127563
- Asher, G., Reuven, N., and Shaul, Y. (2006). 20S proteasomes and protein degradation “by default”. *BioEssays* 28, 844–849. doi: 10.1002/bies.20447
- Babu, M., Bundalovic-Torma, C., Calmettes, C., Phanse, S., Zhang, Q., Jiang, Y., et al. (2018). Global landscape of cell envelope protein complexes in *Escherichia coli*. *Nat. Biotechnol.* 36, 103–112. doi: 10.1038/nbt.4024
- Babu, M., Vlasblom, J., Pu, S., Guo, X., Graham, C., Bean, B. D., et al. (2012). Interaction landscape of membrane-protein complexes in *Saccharomyces cerevisiae*. *Nature* 489, 585–589. doi: 10.1038/nature11354
- Barreto, F. S., Moy, G. W., and Burton, R. S. (2011). Interpopulation patterns of divergence and selection across the transcriptome of the copepod *Tigriopus californicus*. *Mol. Ecol.* 20, 560–572. doi: 10.1111/j.1365-294X.2010.04963.x
- Benschop, J. J., Brabers, N., Van Leenen, D., Bakker, L. V., Van Deutekom, H. W., Van Berkum, N. L., et al. (2010). A consensus of core protein complex compositions for *Saccharomyces cerevisiae*. *Mol. Cell* 38, 916–928. doi: 10.1016/j.molcel.2010.06.002
- Bloom-Ackermann, Z., Navon, S., Gingold, H., Towers, R., Pilpel, Y., and Dahan, O. (2014). A comprehensive tRNA deletion library unravels the genetic architecture of the tRNA pool. *PLoS Genet.* 10:e1004084. doi: 10.1371/journal.pgen.1004084
- Bombliès, K., Lempe, J., Epple, P., Warthmann, N., Lanz, C., Dangel, J. L., et al. (2007). Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLoS Biol.* 5:e236. doi: 10.1371/journal.pbio.0050236
- Bontinck, M., Van Leene, J., Gadeyne, A., De Rybel, B., Eeckhout, D., Nelissen, H., et al. (2018). Recent trends in plant protein complex analysis in a developmental context. *Front. Plant Sci.* 9:640. doi: 10.3389/fpls.2018.00640
- Brideau, N. J., and Barbash, D. A. (2011). Functional conservation of the *Drosophila* hybrid incompatibility gene Lhr. *BMC Evol. Biol.* 11:57. doi: 10.1186/1471-2148-11-57
- Brideau, N. J., Flores, H. A., Wang, J., Maheshwari, S., Wang, X., and Barbash, D. A. (2006). Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *Science* 314, 1292–1295. doi: 10.1126/science.1133953
- Brodie, E. D. 3rd, and Ridenhour, B. J. (2003). Reciprocal selection at the phenotypic interface of coevolution. *Integr. Comp. Biol.* 43, 408–418. doi: 10.1093/icb/43.3.408
- Burger, L., and Van Nimwegen, E. (2010). Disentangling direct from indirect co-evolution of residues in protein alignments. *PLoS Comput. Biol.* 6:e1000633. doi: 10.1371/journal.pcbi.1000633
- Burkart-Waco, D., Josefsson, C., Dilkes, B., Kozloff, N., Torjek, O., Meyer, R., et al. (2012). Hybrid incompatibility in *Arabidopsis* is determined by a multiple-locus genetic network. *Plant Physiol.* 158, 801–812. doi: 10.1104/pp.111.188706
- Burton, R. S., Pereira, R. J., and Barreto, F. S. (2013). Cytonuclear genomic interactions and hybrid breakdown. *Annu. Rev. Ecol. Syst.* 44, 281–302. doi: 10.1146/annurev-ecolsys-110512-135758
- Carroll, S. B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134, 25–36. doi: 10.1016/j.cell.2008.06.030
- Castillo-Davis, C. I., Kondrashov, F. A., Hartl, D. L., and Kulathinal, R. J. (2004). The functional genomic distribution of protein divergence in two animal phyla: coevolution, genomic conflict, and constraint. *Genome Res.* 14, 802–811. doi: 10.1101/gr.2195604
- Chang, A. S., and Noor, M. A. (2007). The genetics of hybrid male sterility between the allopatric species pair *Drosophila persimilis* and *D. pseudoobscura bogotana*: dominant sterility alleles in collinear autosomal regions. *Genetics* 176, 343–349. doi: 10.1534/genetics.106.067314
- Chavez, J. D., Mohr, J. P., Mathay, M., Zhong, X., Keller, A., and Bruce, J. E. (2019). Systems structural biology measurements by in vivo cross-linking with mass spectrometry. *Nat. Protoc.* 14, 2318–2343. doi: 10.1038/s41596-019-0181-3
- Chen, C., Chen, H., Lin, Y. S., Shen, J. B., Shan, J. X., Qi, P., et al. (2014). A two-locus interaction causes interspecific hybrid weakness in rice. *Nat. Commun.* 5:3357. doi: 10.1038/ncomms4357

- Chen, C., Chen, H., Shan, J. X., Zhu, M. Z., Shi, M., Gao, J. P., et al. (2013). Genetic and physiological analysis of a novel type of interspecific hybrid weakness in rice. *Mol. Plant* 6, 716–728. doi: 10.1093/mp/sss146
- Chen, Z., Fischer, L., Tahir, S., Bukowski-Wills, J. C., Barlow, P., and Rappsilber, J. (2016b). Quantitative cross-linking/mass spectrometry reveals subtle protein conformational changes. *Wellcome Open Res.* 1:5. doi: 10.12688/wellcomeopenres.9896.1
- Chen, Z. A., and Rappsilber, J. (2019). Quantitative cross-linking/mass spectrometry to elucidate structural changes in proteins and their complexes. *Nat. Protoc.* 14, 171–201. doi: 10.1038/s41596-018-0089-3
- Chen, X., and Zhang, J. (2012). The ortholog conjecture is untestable by the current gene ontology but is supported by RNA sequencing data. *PLoS Comput. Biol.* 8:e1002784. doi: 10.1371/journal.pcbi.1002784
- Chen, C., Zhiguo, E., and Lin, H. X. (2016a). Evolution and molecular control of hybrid incompatibility in plants. *Front. Plant Sci.* 7:1208. doi: 10.3389/fpls.2016.01208
- Choi, S. S., and Hannehalli, S. (2013). Three independent determinants of protein evolutionary rate. *J. Mol. Evol.* 76, 98–111. doi: 10.1007/s00239-013-9543-6
- Chou, J. Y., Hung, Y. S., Lin, K. H., Lee, H. Y., and Leu, J. Y. (2010). Multiple molecular mechanisms cause reproductive isolation between three yeast species. *PLoS Biol.* 8:e1000432. doi: 10.1371/journal.pbio.1000432
- Chou, J. Y., and Leu, J. Y. (2010). Speciation through cytonuclear incompatibility: insights from yeast and implications for higher eukaryotes. *Bioessays* 32, 401–411. doi: 10.1002/bies.200900162
- Clark, N. L., Gasper, J., Sekino, M., Springer, S. A., Aquadro, C. F., and Swanson, W. J. (2009). Coevolution of interacting fertilization proteins. *PLoS Genet.* 5:e1000570. doi: 10.1371/journal.pgen.1000570
- Collins, S. R., Kemmeren, P., Zhao, X. C., Greenblatt, J. F., Spencer, F., Holstege, F. C., et al. (2007). Toward a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. *Mol. Cell. Proteomics* 6, 439–450. doi: 10.1074/mcp.M600381-MCP200
- Cong, Q., Anishchenko, I., Ovchinnikov, S., and Baker, D. (2019). Protein interaction networks revealed by proteome coevolution. *Science* 365, 185–189. doi: 10.1126/science.aaw6718
- Corbett-Detig, R. B., Zhou, J., Clark, A. G., Hartl, D. L., and Ayroles, J. F. (2013). Genetic incompatibilities are widespread within species. *Nature* 504, 135–137. doi: 10.1038/nature12678
- Costanzo, M., Vandersluis, B., Koch, E. N., Baryshnikova, A., Pons, C., Tan, G., et al. (2016). A global genetic interaction network maps a wiring diagram of cellular function. *Science* 353:aaf1420. doi: 10.1126/science.aaf1420
- Cox, J., Matic, I., Hilger, M., Nagaraj, N., Selbach, M., Olsen, J. V., et al. (2009). A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nat. Protoc.* 4, 698–705. doi: 10.1038/nprot.2009.36
- Coyne, J. A., and Orr, H. A. (2004). Speciation Sinauer Associates. Sunderland, MA 276, 281.
- Das, M., Haberer, G., Panda, A., Das Laha, S., Ghosh, T. C., and Schaffner, A. R. (2016). Expression pattern similarities support the prediction of orthologs retaining common functions after gene duplication events. *Plant Physiol.* 171, 2343–2357. doi: 10.1104/pp.15.01207
- Davies, B., Hatton, E., Altemose, N., Hussin, J. G., Pratto, F., Zhang, G., et al. (2016). Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. *Nature* 530, 171–176. doi: 10.1038/nature16931
- De Wit, P., Pespeni, M. H., Ladner, J. T., Barshis, D. J., Seneca, F., Jaris, H., et al. (2012). The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Mol. Ecol. Resour.* 12, 1058–1067. doi: 10.1111/1755-0998.12003
- Deutsch, E. W., Mendoza, L., Shteynberg, D., Farrah, T., Lam, H., Tasman, N., et al. (2010). A guided tour of the trans-proteomic pipeline. *Proteomics* 10, 1150–1159. doi: 10.1002/pmic.200900375
- Dobson, C. M. (2003). Protein folding and misfolding. *Nature* 426, 884–890. doi: 10.1038/nature02261
- Dobzhansky, T. (1936). Studies on hybrid sterility. II. localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics* 21, 113–135.
- Doerr, A. (2012). Interactomes by mass spectrometry. *Nat. Methods* 9:1043. doi: 10.1038/nmeth.2235
- Dori-Bachash, M., Shalem, O., Manor, Y. S., Pilpel, Y., and Tirosch, I. (2012). Widespread promoter-mediated coordination of transcription and mRNA degradation. *Genome Biol.* 13:R114. doi: 10.1186/gb-2012-13-12-r114
- Dos Reis, M., Savva, R., and Wernisch, L. (2004). Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.* 32, 5036–5044. doi: 10.1093/nar/gkh834
- Drew, K., Lee, C., Huizar, R. L., Tu, F., Borgeson, B., Mcwhite, C. D., et al. (2017a). Integration of over 9,000 mass spectrometry experiments builds a global map of human protein complexes. *Mol. Syst. Biol.* 13:932. doi: 10.15252/msb.20167490
- Drew, K., Muller, C. L., Bonneau, R., and Marcotte, E. M. (2017b). Identifying direct contacts between protein complex subunits from their conditional dependence in proteomics datasets. *PLoS Comput. Biol.* 13:e1005625. doi: 10.1371/journal.pcbi.1005625
- Dunker, A. K., Silman, I., Uversky, V. N., and Sussman, J. L. (2008). Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* 18, 756–764. doi: 10.1016/j.sbi.2008.10.002
- El Khamlichi, C., Reverchon-Assadi, F., Hervouet-Coste, N., Blot, L., Reiter, E., and Morisset-Lopez, S. (2019). Bioluminescence resonance energy transfer as a method to study protein-protein interactions: application to G protein coupled receptor biology. *Molecules* 24:537. doi: 10.3390/molecules24030537
- Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends Ecol. Evol.* 29, 51–63. doi: 10.1016/j.tree.2013.09.008
- Ellis, R. J. (2007). Protein misassembly: macromolecular crowding and molecular chaperones. *Adv. Exp. Med. Biol.* 594, 1–13. doi: 10.1007/978-0-387-39975-1_1
- Enright, A. J., Van Dongen, S., and Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30, 1575–1584. doi: 10.1093/nar/30.7.1575
- Fraisse, C., Gunnarsson, P. A., Roze, D., Bierne, N., and Welch, J. J. (2016). The genetics of speciation: insights from Fisher's geometric model. *Evolution* 70, 1450–1464. doi: 10.1111/evo.12968
- Francis, B. R., and Thorsness, P. E. (2011). Hsp90 and mitochondrial proteases Yme1 and Yta10/12 participate in ATP synthase assembly in *Saccharomyces cerevisiae*. *Mitochondrion* 11, 587–600. doi: 10.1016/j.mito.2011.03.008
- Fraser, H. B. (2005). Modularity and evolutionary constraint on proteins. *Nat. Genet.* 37, 351–352. doi: 10.1038/ng1530
- Fraser, H. B., and Hirsh, A. E. (2004). Evolutionary rate depends on number of protein-protein interactions independently of gene expression level. *BMC Evol. Biol.* 4:13. doi: 10.1186/1471-2148-4-13
- Fraser, H. B., Hirsh, A. E., Wall, D. P., and Eisen, M. B. (2004). Coevolution of gene expression among interacting proteins. *Proc. Natl. Acad. Sci. U. S. A.* 101, 9033–9038. doi: 10.1073/pnas.0402591101
- Fraser, H. B., Wall, D. P., and Hirsh, A. E. (2003). A simple dependence between protein evolution rate and the number of protein-protein interactions. *BMC Evol. Biol.* 3:11. doi: 10.1186/1471-2148-3-11
- Gagnaire, P. A., Normandeau, E., and Bernatchez, L. (2012). Comparative genomics reveals adaptive protein evolution and a possible cytonuclear incompatibility between European and American Eels. *Mol. Biol. Evol.* 29, 2909–2919. doi: 10.1093/molbev/mss076
- Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., et al. (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature* 440, 631–636. doi: 10.1038/nature04532
- Ge, H., Liu, Z., Church, G. M., and Vidal, M. (2001). Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nat. Genet.* 29, 482–486. doi: 10.1038/ng776
- Giani, A. M., Gallo, G. R., Gianfranceschi, L., and Formenti, G. (2020). Long walk to genomics: history and current approaches to genome sequencing and assembly. *Comput. Struct. Biotechnol. J.* 18, 9–19. doi: 10.1016/j.csbj.2019.11.002
- Gidalevitz, T., Prahlad, V., and Morimoto, R. I. (2011). The stress of protein misfolding: from single cells to multicellular organisms. *Cold Spring Harb. Perspect. Biol.* 3:a009704. doi: 10.1101/cshperspect.a009704
- Gilad, Y., Oshlack, A., and Rifkin, S. A. (2006). Natural selection on gene expression. *Trends Genet.* 22, 456–461. doi: 10.1016/j.tig.2006.06.002
- Gilad, Y., Pritchard, J. K., and Thornton, K. (2009). Characterizing natural variation using next-generation sequencing technologies. *Trends Genet.* 25, 463–471. doi: 10.1016/j.tig.2009.09.003
- Gloge, F., Becker, A. H., Kramer, G., and Bukau, B. (2014). Co-translational mechanisms of protein maturation. *Curr. Opin. Struct. Biol.* 24, 24–33. doi: 10.1016/j.sbi.2013.11.004
- Gopinath, R. K., You, S. T., Chien, K. Y., Swamy, K. B., Yu, J. S., Schuyler, S. C., et al. (2014). The Hsp90-dependent proteome is conserved and enriched

- for hub proteins with high levels of protein-protein connectivity. *Genome Biol. Evol.* 6, 2851–2865. doi: 10.1093/gbe/evu226
- Gorka, M., Swart, C., Siemiakowska, B., Martinez-Jaime, S., Skiryycz, A., Streb, S., et al. (2019). Protein complex identification and quantitative complexome by CN-PAGE. *Sci. Rep.* 9:11523. doi: 10.1038/s41598-019-47829-7
- Guruharsha, K. G., Rual, J. F., Zhai, B., Mintseris, J., Vaidya, P., Vaidya, N., et al. (2011). A protein complex network of *Drosophila melanogaster*. *Cell* 147, 690–703. doi: 10.1016/j.cell.2011.08.047
- Harrison, J. S., and Burton, R. S. (2006). Tracing hybrid incompatibilities to single amino acid substitutions. *Mol. Biol. Evol.* 23, 559–564. doi: 10.1093/molbev/msj058
- Hart, G. T., Lee, I., and Marcotte, E. R. (2007). A high-accuracy consensus map of yeast protein complexes reveals modular nature of gene essentiality. *BMC Bioinformatics* 8:236. doi: 10.1186/1471-2105-8-236
- Hartl, F. U., Bracher, A., and Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. *Nature* 475, 324–332. doi: 10.1038/nature10317
- Havugimana, P. C., Hart, G. T., Nepusz, T., Yang, H., Turinsky, A. L., Li, Z., et al. (2012). A census of human soluble protein complexes. *Cell* 150, 1068–1081. doi: 10.1016/j.cell.2012.08.011
- Hawe, J. S., Theis, F. J., and Heinig, M. (2019). Inferring interaction networks from multi-omics data. *Front. Genet.* 10:535. doi: 10.3389/fgenet.2019.00535
- Henault, M., and Landry, C. R. (2017). When nuclear-encoded proteins and mitochondrial RNAs do not get along, species split apart. *EMBO Rep.* 18, 8–10. doi: 10.15252/embr.201643645
- Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., et al. (2002). Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415, 180–183. doi: 10.1038/415180a
- Hodgins-Davis, A., Rice, D. P., and Townsend, J. P. (2015). Gene expression evolves under a house-of-cards model of stabilizing selection. *Mol. Biol. Evol.* 32, 2130–2140. doi: 10.1093/molbev/msv094
- Hoekstra, H. E., and Coyne, J. A. (2007). The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 61, 995–1016. doi: 10.1111/j.1558-5646.2007.00105.x
- Hopf, T. A., Scharfe, C. P., Rodrigues, J. P., Green, A. G., Kohlbacher, O., Sander, C., et al. (2014). Sequence co-evolution gives 3D contacts and structures of protein complexes. *Elife* 3:e03430. doi: 10.7554/eLife.03430
- Hou, J., Fournier, T., and Schacherer, J. (2016). Species-wide survey reveals the various flavors of intraspecific reproductive isolation in yeast. *FEMS Yeast Res.* 16:fow048. doi: 10.1093/femsyr/fow048
- Hou, J., Friedrich, A., Gounot, J. S., and Schacherer, J. (2015). Comprehensive survey of condition-specific reproductive isolation reveals genetic incompatibility in yeast. *Nat. Commun.* 6:7214. doi: 10.1038/ncomms8214
- Hu, L. Z., Goebels, E., Tan, J. H., Wolf, E., Kuzmanov, U., Wan, C., et al. (2019). EPIC: software toolkit for elution profile-based inference of protein complexes. *Nat. Methods* 16, 737–742. doi: 10.1038/s41592-019-0461-4
- Hua, J. (2013). Modulation of plant immunity by light, circadian rhythm, and temperature. *Curr. Opin. Plant Biol.* 16, 406–413. doi: 10.1016/j.pbi.2013.06.017
- Jansen, R., Greenbaum, D., and Gerstein, M. (2002). Relating whole-genome expression data with protein-protein interactions. *Genome Res.* 12, 37–46. doi: 10.1101/gr.205602
- Jhuang, H. Y., Lee, H. Y., and Leu, J. Y. (2017). Mitochondrial-nuclear co-evolution leads to hybrid incompatibility through pentatricopeptide repeat proteins. *EMBO Rep.* 18, 87–101. doi: 10.15252/embr.201643311
- Johnson, N. A. (2010). Hybrid incompatibility genes: remnants of a genomic battlefield? *Trends Genet.* 26, 317–325. doi: 10.1016/j.tig.2010.04.005
- Juan, D., Pazos, F., and Valencia, A. (2008). Co-evolution and co-adaptation in protein networks. *FEBS Lett.* 582, 1225–1230. doi: 10.1016/j.febslet.2008.02.017
- Kalirad, A., and Azevedo, R. B. R. (2017). Spiraling complexity: a test of the snowball effect in a computational model of RNA folding. *Genetics* 206, 377–388. doi: 10.1534/genetics.116.196030
- Kao, K. C., Schwartz, K., and Sherlock, G. (2010). A genome-wide analysis reveals no nuclear dobzhansky-muller pairs of determinants of speciation between *S. cerevisiae* and *S. paradoxus*, but suggests more complex incompatibilities. *PLoS Genet.* 6:e1001038. doi: 10.1371/journal.pgen.1001038
- Kastritis, P. L., O'Reilly, F. J., Bock, T., Li, Y., Rogon, M. Z., Buczak, K., et al. (2017). Capturing protein communities by structural proteomics in a thermophilic eukaryote. *Mol. Syst. Biol.* 13:936. doi: 10.15252/msb.20167412
- Klink, G. V., and Bazykin, G. A. (2017). Parallel evolution of metazoan mitochondrial proteins. *Genome Biol. Evol.* 9, 1341–1350. doi: 10.1093/gbe/evx025
- Kohlbacher, O., Reinert, K., Gropl, C., Lange, E., Pfeifer, N., Schulz-Trieglaff, O., et al. (2007). TOPP—the OpenMS proteomics pipeline. *Bioinformatics* 23, e191–e197. doi: 10.1093/bioinformatics/btl299
- Kondrashov, A. S. (2003). Accumulation of Dobzhansky-Muller incompatibilities within a spatially structured population. *Evolution* 57, 151–153. doi: 10.1111/j.0014-3820.2003.tb00223.x
- Kristensen, A. R., Gsponer, J., and Foster, L. J. (2012). A high-throughput approach for measuring temporal changes in the interactome. *Nat. Methods* 9, 907–909. doi: 10.1038/nmeth.2131
- Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., et al. (2006). Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 440, 637–643. doi: 10.1038/nature04670
- Lachance, J., and Tishkoff, S. A. (2013). SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it. *Bioessays* 35, 780–786. doi: 10.1002/bies.201300014
- Lai, H. Y., Yu, Y. H., and Leu, J. Y. (2018). Multiple intermolecular interactions facilitate rapid evolution of essential genes. *SSRN Electron. J.* doi: 10.2139/ssrn.3155896 [Preprint]
- Landguth, E. L., Johnson, N. A., and Cushman, S. A. (2015). Clusters of incompatible genotypes evolve with limited dispersal. *Front. Genet.* 6:151. doi: 10.3389/fgenet.2015.00151
- Landry, C. R., Wittkopp, P. J., Taubes, C. H., Ranz, J. M., Clark, A. G., and Hartl, D. L. (2005). Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of *Drosophila*. *Genetics* 171, 1813–1822. doi: 10.1534/genetics.105.047449
- Leducq, J. B., Charron, G., Diss, G., Gagnon-Arsenault, I., Dube, A. K., and Landry, C. R. (2012). Evidence for the robustness of protein complexes to inter-species hybridization. *PLoS Genet.* 8:e1003161. doi: 10.1371/journal.pgen.1003161
- Lee, H. Y., Chou, J. Y., Cheong, L., Chang, N. H., Yang, S. Y., and Leu, J. Y. (2008). Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell* 135, 1065–1073. doi: 10.1016/j.cell.2008.10.047
- Lee, S., Lee, H., Yoo, S., and Kim, H. (2017). Molecular insights into the m-AAA protease-mediated dislocation of transmembrane helices in the mitochondrial inner membrane. *J. Biol. Chem.* 292, 20058–20066. doi: 10.1074/jbc.M117.796763
- Lee, Y. W., Gould, B. A., and Stinchcombe, J. R. (2014). Identifying the genes underlying quantitative traits: a rationale for the QTN programme. *AoB Plants* 6:plu004. doi: 10.1093/aobpla/plu004
- Levinthal, C. (1969). “How to fold graciously” in *Mossbauer spectroscopy in biological systems: Proceedings of a meeting held at Allerton house*. eds. J. T. P. Debrunner and E. Munck (Champaign, IL: University of Illinois Press), 22–24.
- Li, C., Wang, Z., and Zhang, J. (2013). Toward genome-wide identification of Bateson-Dobzhansky-Muller incompatibilities in yeast: a simulation study. *Genome Biol. Evol.* 5, 1261–1272. doi: 10.1093/gbe/evt091
- Li, J., Soroka, J., and Buchner, J. (2012). The Hsp90 chaperone machinery: conformational dynamics and regulation by co-chaperones. *Biochim. Biophys. Acta* 1823, 624–635. doi: 10.1016/j.bbamcr.2011.09.003
- Liao, X., Li, M., Zou, Y., Wu, F. X., Pan, Y., and Wang, J. (2019). Current challenges and solutions of de novo assembly. *Quant. Biol.* 7, 90–109. doi: 10.1007/s40484-019-0166-9
- Lim, Y., Yu, I., Seo, D., Kang, U., and Sael, L. (2019). PS-MCL: parallel shotgun coarsened Markov clustering of protein interaction networks. *BMC Bioinformatics* 20:381. doi: 10.1186/s12859-019-2856-8
- Lindtke, D., and Buerkle, C. A. (2015). The genetic architecture of hybrid incompatibilities and their effect on barriers to introgression in secondary contact. *Evolution* 69, 1987–2004. doi: 10.1111/evo.12725
- Lopez, P., Casane, D., and Philippe, H. (2002). Heterotachy, an important process of protein evolution. *Mol. Biol. Evol.* 19, 1–7. doi: 10.1093/oxfordjournals.molbev.a003973
- Lopez-Fernandez, H., and Bolnick, D. I. (2007). What causes partial F1 hybrid viability? Incomplete penetrance versus genetic variation. *PLoS One* 2:e1294. doi: 10.1371/journal.pone.0001294
- Mack, K. L., and Nachman, M. W. (2017). Gene regulation and speciation. *Trends Genet.* 33, 68–80. doi: 10.1016/j.tig.2016.11.003
- Maheshwari, S., and Barbash, D. A. (2011). The genetics of hybrid incompatibilities. *Annu. Rev. Genet.* 45, 331–355. doi: 10.1146/annurev-genet-110410-132514

- Makhnevych, T., and Houry, W. A. (2012). The role of Hsp90 in protein complex assembly. *Biochim. Biophys. Acta* 1823, 674–682. doi: 10.1016/j.bbamcr.2011.09.001
- Malty, R. H., Aoki, H., Kumar, A., Phanse, S., Amin, S., Zhang, Q., et al. (2017). A map of human mitochondrial protein interactions linked to neurodegeneration reveals new mechanisms of redox homeostasis and NF-kappaB signaling. *Cell Syst* 5, 564–577. doi: 10.1016/j.cels.2017.10.010
- Marks, D. S., Hopf, T. A., and Sander, C. (2012). Protein structure prediction from sequence variation. *Nat. Biotechnol.* 30, 1072–1080. doi: 10.1038/nbt.2419
- Marsh, J. A., Hernandez, H., Hall, Z., Ahnert, S. E., Perica, T., Robinson, C. V., et al. (2013). Protein complexes are under evolutionary selection to assemble via ordered pathways. *Cell* 153, 461–470. doi: 10.1016/j.cell.2013.02.044
- Marsh, J. A., and Teichmann, S. A. (2015). Structure, dynamics, assembly, and evolution of protein complexes. *Annu. Rev. Biochem.* 84, 551–575. doi: 10.1146/annurev-biochem-060614-034142
- Mashaghi, A., Bezrukavnikov, S., Minde, D. P., Wentink, A. S., Kityk, R., Zachmann-Brand, B., et al. (2016). Alternative modes of client binding enable functional plasticity of Hsp70. *Nature* 539, 448–451. doi: 10.1038/nature20137
- Matute, D. R., Butler, I. A., Turissini, D. A., and Coyne, J. A. (2010). A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science* 329, 1518–1521. doi: 10.1126/science.1193440
- McClellan, A. J., Xia, Y., Deutschbauer, A. M., Davis, R. W., Gerstein, M., and Frydman, J. (2007). Diverse cellular functions of the Hsp90 molecular chaperone uncovered using systems approaches. *Cell* 131, 121–135. doi: 10.1016/j.cell.2007.07.036
- McGirr, J. A., and Martin, C. H. (2019). Hybrid gene misregulation in multiple developing tissues within a recent adaptive radiation of Cyprinodon pupfishes. *PLoS One* 14:e0218899. doi: 10.1371/journal.pone.0218899
- McWhite, C. D., Papoulas, O., Drew, K., Cox, R. M., June, V., Dong, O. X., et al. (2020). A pan-plant protein complex map reveals deep conservation and novel assemblies. *Cell* 181:e414. doi: 10.1016/j.cell.2020.02.049
- Meiklejohn, C. D., Holmbeck, M. A., Siddiq, M. A., Abt, D. N., Rand, D. M., and Montooth, K. L. (2013). An incompatibility between a mitochondrial tRNA and its nuclear-encoded tRNA synthetase compromises development and fitness in *Drosophila*. *PLoS Genet.* 9:e1003238. doi: 10.1371/journal.pgen.1003238
- Michnick, S. W., Ear, P. H., Landry, C., Malleshaiah, M. K., and Messier, V. (2010). A toolkit of protein-fragment complementation assays for studying and dissecting large-scale and dynamic protein-protein interactions in living cells. *Methods Enzymol.* 470, 335–368. doi: 10.1016/S0076-6879(10)70014-8
- Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J. C., and Forejt, J. (2009). A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* 323, 373–375. doi: 10.1126/science.1163601
- Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D. S., Sander, C., et al. (2011). Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci. U. S. A.* 108, E1293–E1301. doi: 10.1073/pnas.1111471108
- Moyle, L. C., and Nakazato, T. (2009). Complex epistasis for Dobzhansky-Muller hybrid incompatibility in *solanum*. *Genetics* 181, 347–351. doi: 10.1534/genetics.108.095679
- Muller, H. J. (1939). Reversibility in evolution considered from the standpoint of genetics. *Biol. Rev.* 14, 261–280. doi: 10.1111/j.1469-185X.1939.tb00934.x
- Nahnsen, S., Bertsch, A., Rahnenfuhrer, J., Nordheim, A., and Kohlbacher, O. (2011). Probabilistic consensus scoring improves tandem mass spectrometry peptide identification. *J. Proteome Res.* 10, 3332–3343. doi: 10.1021/pr2002879
- Nesvizhskii, A. I. (2014). Proteogenomics: concepts, applications and computational strategies. *Nat. Methods* 11, 1114–1125. doi: 10.1038/nmeth.3144
- Nesvizhskii, A. I., Vitek, O., and Aebersold, R. (2007). Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat. Methods* 4, 787–797. doi: 10.1038/nmeth1088
- Neverov, A. D., Fedonin, G. G., Cheremukhin, E. A., Klink, G. V., Popova, A. V., and Bazykin, G. A. (2020). Episodic evolution of coadapted sets of amino acid sites in mitochondrial proteins. bioRxiv [Preprint].
- Nusinow, D. P., Szpyt, J., Ghandi, M., Rose, C. M., McDonald, E. R. 3rd, Kalocsy, M., et al. (2020). Quantitative proteomics of the cancer cell line encyclopaedia. *Cell* 180:e316. doi: 10.1016/j.cell.2019.12.023
- Ochoa, D., and Pazos, F. (2014). Practical aspects of protein co-evolution. *Front. Cell Dev. Biol.* 2:14. doi: 10.3389/fcell.2014.00014
- Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., et al. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* 1, 376–386. doi: 10.1074/mcp.M200025-MCP200
- Oromendia, A. B., and Amon, A. (2014). Aneuploidy: implications for protein homeostasis and disease. *Dis. Model. Mech.* 7, 15–20. doi: 10.1242/dmm.013391
- Orr, H. A. (1995). The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* 139, 1805–1813. doi: 10.1093/genetics/139.4.1805
- Orr, H. A., and Turelli, M. (2001). The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution* 55, 1085–1094. doi: 10.1111/j.0014-3820.2001.tb00628.x
- Oshlack, A., Robinson, M. D., and Young, M. D. (2010). From RNA-seq reads to differential expression results. *Genome Biol.* 11:220. doi: 10.1186/gb-2010-11-12-220
- Paiano, A., Margiotta, A., De Luca, M., and Bucci, C. (2019). Yeast two-hybrid assay to identify interacting proteins. *Curr. Protoc. Protein Sci.* 95:e70. doi: 10.1002/cpps.70
- Paliwal, S., Fiumera, A. C., and Fiumera, H. L. (2014). Mitochondrial-nuclear epistasis contributes to phenotypic variation and coadaptation in natural isolates of *Saccharomyces cerevisiae*. *Genetics* 198, 1251–1265. doi: 10.1534/genetics.114.168575
- Phadnis, N., Baker, E. P., Cooper, J. C., Frizzell, K. A., Hsieh, E., De La Cruz, A. F., et al. (2015). An essential cell cycle regulation gene causes hybrid inviability in *Drosophila*. *Science* 350, 1552–1555. doi: 10.1126/science.aac7504
- Piatkowska, E. M., Naseeb, S., Knight, D., and Delneri, D. (2013). Chimeric protein complexes in hybrid species generate novel phenotypes. *PLoS Genet.* 9:e1003836. doi: 10.1371/journal.pgen.1003836
- Presgraves, D. C. (2010a). The molecular evolutionary basis of species formation. *Nat. Rev. Genet.* 11, 175–180. doi: 10.1038/nrg2718
- Presgraves, D. C. (2010b). Speciation genetics: search for the missing snowball. *Curr. Biol.* 20, R1073–R1074. doi: 10.1016/j.cub.2010.10.056
- Price, T., Pena, F. I. 3rd, and Cho, Y. R. (2013). Survey: enhancing protein complex prediction in PPI networks with GO similarity weighting. *Interdiscip. Sci.* 5, 196–210. doi: 10.1007/s12539-013-0174-9
- Pu, S., Wong, J., Turner, B., Cho, E., and Wodak, S. J. (2009). Up-to-date catalogues of yeast protein complexes. *Nucleic Acids Res.* 37, 825–831. doi: 10.1093/nar/gkn1005
- Qian, W., He, X., Chan, E., Xu, H., and Zhang, J. (2011). Measuring the evolutionary rate of protein-protein interaction. *Proc. Natl. Acad. Sci. U. S. A.* 108, 8725–8730. doi: 10.1073/pnas.1104695108
- Radwan, M., Wood, R. J., Sui, X., and Hatters, D. M. (2017). When proteostasis goes bad: protein aggregation in the cell. *IUBMB Life* 69, 49–54. doi: 10.1002/iub.1597
- Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlovic, M., et al. (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J. Evol. Biol.* 30, 1450–1477. doi: 10.1111/jeb.13047
- Rieseberg, L. H., and Blackman, B. K. (2010). Speciation genes in plants. *Ann. Bot.* 106, 439–455. doi: 10.1093/aob/mcq126
- Rizzato, F., Zamuner, S., Pagnani, A., and Laio, A. (2019). A common root for coevolution and substitution rate variability in protein sequence evolution. *Sci. Rep.* 9:18032. doi: 10.1038/s41598-019-53958-w
- Rodriguez-Cruz, M., Coral-Vazquez, R. M., Hernandez-Stengele, G., Sanchez, R., Salazar, E., Sanchez-Munoz, F., et al. (2013). Identification of putative ortholog gene blocks involved in gestant and lactating mammary gland development: a rodent cross-species microarray transcriptomics approach. *Int. J. Genomics* 2013:624681. doi: 10.1155/2013/624681
- Roure, B., and Philippe, H. (2011). Site-specific time heterogeneity of the substitution process and its impact on phylogenetic inference. *BMC Evol. Biol.* 11:17. doi: 10.1186/1471-2148-11-17
- Ruggles, K. V., Krug, K., Wang, X., Clauser, K. R., Wang, J., Payne, S. H., et al. (2017). Methods, tools and current perspectives in proteogenomics. *Mol. Cell. Proteomics* 16, 959–981. doi: 10.1074/mcp.MR117.000024
- Satuluri, V., and Parthasarathy, S. (2009). “Scalable graph clustering using stochastic flows: Applications to community discovery,” in *The 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining: ACM*; June 2009; 737–746.

- Satuluri, V., Parthasarathy, S., and Ucar, D. (2010). "Markov clustering of protein interaction networks with improved balance and scalability." in *ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, New York; August 2010.
- Schoenrock, A., Burnside, D., Moteshareie, H., Pitre, S., Hooshyar, M., Green, J. R., et al. (2017). Evolution of protein-protein interaction networks in yeast. *PLoS One* 12:e0171920. doi: 10.1371/journal.pone.0171920
- Schumer, M., Cui, R., Rosenthal, G. G., and Andolfatto, P. (2015). Reproductive isolation of hybrid populations driven by genetic incompatibilities. *PLoS Genet.* 11:e1005041. doi: 10.1371/journal.pgen.1005041
- Schwarz, A., and Beck, M. (2019). The benefits of cotranslational assembly: a structural perspective. *Trends Cell Biol.* 29, 791–803. doi: 10.1016/j.tcb.2019.07.006
- Shiber, A., Doring, K., Friedrich, U., Klann, K., Merker, D., Zedan, M., et al. (2018). Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling. *Nature* 561, 268–272. doi: 10.1038/s41586-018-0462-y
- Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Sinz, A. (2006). Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom. Rev.* 25, 663–682. doi: 10.1002/mas.20082
- Song, J., Pfanner, N., and Becker, T. (2018). Assembling the mitochondrial ATP synthase. *Proc. Natl. Acad. Sci. U. S. A.* 115, 2850–2852. doi: 10.1073/pnas.1801697115
- Srihari, S., and Leong, H. W. (2012). Employing functional interactions for characterisation and detection of sparse complexes from yeast PPI networks. *Int. J. Bioinforma. Res. Appl.* 8, 286–304. doi: 10.1504/IJBRA.2012.048962
- Srihari, S., and Leong, H. W. (2013). A survey of computational methods for protein complex prediction from protein interaction networks. *J. Bioinforma. Comput. Biol.* 11:1230002. doi: 10.1142/S021972001230002X
- Stinchcombe, J. R., and Hoekstra, H. E. (2008). Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 100, 158–170. doi: 10.1038/sj.hdy.6800937
- Sudhir, P. R., and Chen, C. H. (2016). Proteomics-based analysis of protein complexes in pluripotent stem cells and cancer biology. *Int. J. Mol. Sci.* 17:432. doi: 10.3390/ijms17030432
- Sun, J., Wang, M., Wang, H., Zhang, H., Zhang, X., Thiagarajan, V., et al. (2012). De novo assembly of the transcriptome of an invasive snail and its multiple ecological applications. *Mol. Ecol. Resour.* 12, 1133–1144. doi: 10.1111/1755-0998.12014
- Swain Lenz, D., Riles, L., and Fay, J. C. (2014). Heterochronic meiotic misexpression in an interspecific yeast hybrid. *Mol. Biol. Evol.* 31, 1333–1342. doi: 10.1093/molbev/msu098
- Tang, S., and Presgraves, D. C. (2009). Evolution of the *Drosophila* nuclear pore complex results in multiple hybrid incompatibilities. *Science* 323, 779–782. doi: 10.1126/science.1169123
- Tarassov, K., Messier, V., Landry, C. R., Radinovic, S., Serna Molina, M. M., Shames, I., et al. (2008). An in vivo map of the yeast protein interactome. *Science* 320, 1465–1470. doi: 10.1126/science.1153878
- Telschow, A., Gadau, J., Werren, J. H., and Kobayashi, Y. (2019). Genetic incompatibilities between mitochondria and nuclear genes: effect on gene flow and speciation. *Front. Genet.* 10:62. doi: 10.3389/fgene.2019.00062
- Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., et al. (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* 75, 1895–1904. doi: 10.1021/ac0262560
- Ting, C. T., Tsaui, S. C., Wu, M. L., and Wu, C. I. (1998). A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* 282, 1501–1504. doi: 10.1126/science.282.5393.1501
- Tirosh, I., and Barkai, N. (2011). Inferring regulatory mechanisms from patterns of evolutionary divergence. *Mol. Syst. Biol.* 7:530. doi: 10.1038/msb.2011.60
- Tirosh, I., Reikhav, S., Levy, A. A., and Barkai, N. (2009). A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* 324, 659–662. doi: 10.1126/science.1169766
- Traw, M. B., and Bergelson, J. (2010). Plant immune system incompatibility and the distribution of enemies in natural hybrid zones. *Curr. Opin. Plant Biol.* 13, 466–471. doi: 10.1016/j.pbi.2010.04.009
- True, J. R., and Haag, E. S. (2001). Developmental system drift and flexibility in evolutionary trajectories. *Evol. Dev.* 3, 109–119. doi: 10.1046/j.1525-142x.2001.003002109.x
- Turelli, M., and Moyle, L. C. (2007). Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. *Genetics* 176, 1059–1088. doi: 10.1534/genetics.106.065979
- Turelli, M., and Orr, H. A. (2000). Dominance, epistasis and the genetics of postzygotic isolation. *Genetics* 154, 1663–1679.
- Van Dongen, S., and Abreu-Goodger, C. (2012). Using MCL to extract clusters from networks. *Methods Mol. Biol.* 804, 281–295. doi: 10.1007/978-1-61779-361-5_15
- Volkel, P., Le Faou, P., and Angrand, P. O. (2010). Interaction proteomics: characterization of protein complexes using tandem affinity purification-mass spectrometry. *Biochem. Soc. Trans.* 38, 883–887. doi: 10.1042/BST0380883
- Wade, M. (2002). A gene's eye view of epistasis, selection and speciation. *J. Evol. Biol.* 15, 337–346. doi: 10.1046/j.1420-9101.2002.00413.x
- Wan, C., Borgeson, B., Phanse, S., Tu, F., Drew, K., Clark, G., et al. (2015). Panorama of ancient metazoan macromolecular complexes. *Nature* 525, 339–344. doi: 10.1038/nature14877
- Wang, H., Kakaradov, B., Collins, S. R., Karotki, L., Fiedler, D., Shales, M., et al. (2009a). A complex-based reconstruction of the *Saccharomyces cerevisiae* interactome. *Mol. Cell. Proteomics* 8, 1361–1381. doi: 10.1074/mcp.M800490-MCP200
- Wang, R., Liu, G., and Wang, C. (2019). Identifying protein complexes based on an edge weight algorithm and core-attachment structure. *BMC Bioinformatics* 20:471. doi: 10.1186/s12859-019-3231-5
- Wang, X., Liu, Q., and Zhang, B. (2014). Leveraging the complementary nature of RNA-Seq and shotgun proteomics data. *Proteomics* 14, 2676–2687. doi: 10.1002/pmic.201400184
- Wang, Z., Gerstein, M., and Snyder, M. (2009b). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484
- Wang, Z., Yu, K., Tan, H., Wu, Z., Cho, J. H., Han, X., et al. (2020). 27-plex tandem mass tag mass spectrometry for profiling brain proteome in Alzheimer's disease. *Anal. Chem.* 92, 7162–7170. doi: 10.1021/acs.analchem.0c00655
- Wang, X., and Zhang, B. (2014). Integrating genomic, transcriptomic, and interactome data to improve peptide and protein identification in shotgun proteomics. *J. Proteome Res.* 13, 2715–2723. doi: 10.1021/pr500194t
- Watt, I. N., Montgomery, M. G., Runswick, M. J., Leslie, A. G., and Walker, J. E. (2010). Bioenergetic cost of making an adenosine triphosphate molecule in animal mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* 107, 16823–16827. doi: 10.1073/pnas.1011099107
- Weigt, M., White, R. A., Szurmant, H., Hoch, J. A., and Hwa, T. (2009). Identification of direct residue contacts in protein-protein interaction by message passing. *Proc. Natl. Acad. Sci. U. S. A.* 106, 67–72. doi: 10.1073/pnas.0805923106
- Welch, J. J. (2004). Accumulating Dobzhansky-Muller incompatibilities: reconciling theory and data. *Evolution* 58, 1145–1156. doi: 10.1111/j.0014-3820.2004.tb01695.x
- Will, T., and Helms, V. (2019). Differential analysis of combinatorial protein complexes with ComplexXChange. *BMC Bioinformatics* 20:300. doi: 10.1186/s12859-019-2852-z
- Williams, S. M., Liyu, A. V., Tsai, C. F., Moore, R. J., Orton, D. J., Chrisler, W. B., et al. (2020). Automated coupling of nanodroplet sample preparation with liquid chromatography-mass spectrometry for high-throughput single-cell proteomics. *Anal. Chem.* 92, 10588–10596. doi: 10.1021/acs.analchem.0c01551
- Wittbrodt, J., Adam, D., Malitschek, B., Maueler, W., Raulf, F., Telling, A., et al. (1989). Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu locus in *Xiphophorus*. *Nature* 341, 415–421. doi: 10.1038/341415a0
- Wolf, J. B., and Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nat. Rev. Genet.* 18, 87–100. doi: 10.1038/nrg.2016.133
- Wolf, J. B., Lindell, J., and Backstrom, N. (2010). Speciation genetics: current status and evolving approaches. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 365, 1717–1733. doi: 10.1098/rstb.2010.0023
- Wolters, J. F., Charron, G., Gaspary, A., Landry, C. R., Fiumera, A. C., and Fiumera, H. L. (2018). Mitochondrial recombination reveals mito-mito epistasis in yeast. *Genetics* 209, 307–319. doi: 10.1534/genetics.117.300660
- Wolters, J. F., Chiu, K., and Fiumera, H. L. (2015). Population structure of mitochondrial genomes in *Saccharomyces cerevisiae*. *BMC Genomics* 16:451. doi: 10.1186/s12864-015-1664-4

- Wu, C. I., and Palopoli, M. F. (1994). Genetics of postmating reproductive isolation in animals. *Annu. Rev. Genet.* 28, 283–308. doi: 10.1146/annurev.ge.28.120194.001435
- Yandell, M., and Ence, D. (2012). A beginner's guide to eukaryotic genome annotation. *Nat. Rev. Genet.* 13, 329–342. doi: 10.1038/nrg3174
- Yates, J. R. 3rd, Park, S. K., Delahunty, C. M., Xu, T., Savas, J. N., Cociorva, D., et al. (2012). Toward objective evaluation of proteomic algorithms. *Nat. Methods* 9, 455–456. doi: 10.1038/nmeth.1983
- Yona, A. H., Bloom-Ackermann, Z., Frumkin, I., Hanson-Smith, V., Charpak-Amikam, Y., Feng, Q., et al. (2013). tRNA genes rapidly change in evolution to meet novel translational demands. *Elife* 2:e01339. doi: 10.7554/eLife.01339
- Zahiri, J., Emamjomeh, A., Bagheri, S., Ivazeh, A., Mahdevar, G., Sepasi Tehrani, H., et al. (2020). Protein complex prediction: a survey. *Genomics* 112, 174–183. doi: 10.1016/j.ygeno.2019.01.011
- Zamir, L., Zaretsky, M., Fridman, Y., Ner-Gaon, H., Rubin, E., and Aharoni, A. (2012). Tight coevolution of proliferating cell nuclear antigen (PCNA)-partner interaction networks in fungi leads to interspecies network incompatibility. *Proc. Natl. Acad. Sci. U. S. A.* 109, E406–E414. doi: 10.1073/pnas.1108633109
- Zhang, J., and Gu, X. (1998). Correlation between the substitution rate and rate variation among sites in protein evolution. *Genetics* 149, 1615–1625.
- Zhang, L. V., Wong, S. L., King, O. D., and Roth, F. P. (2004). Predicting co-complexed protein pairs using genomic and proteomic data integration. *BMC Bioinformatics* 5:38. doi: 10.1186/1471-2105-5-89
- Zhong, Q., Pevzner, S. J., Hao, T., Wang, Y., Mosca, R., Menche, J., et al. (2016). An inter-species protein-protein interaction network across vast evolutionary distance. *Mol. Syst. Biol.* 12:865. doi: 10.15252/msb.20156484
- Zill, O. A., Scannell, D. R., Kuei, J., Sadhu, M., and Rine, J. (2012). Evolutionary analysis of heterochromatin protein compatibility by interspecies complementation in *Saccharomyces*. *Genetics* 192, 1001–1014. doi: 10.1534/genetics.112.141549

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Swamy, Schuyler and Leu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.