

Contents

May 2017 • VOLUME 7 • ISSUE 5 • www.g3journal.org

INVESTIGATIONS

- 1385–1392 **Accurate Classification of Protein Subcellular Localization from High-Throughput Microscopy Images Using Deep Learning**
Tanel Pärnamaa and Leopold Parts
- Deep learning, an exciting class of analysis methods, is applied to data from high throughput microscopy screens. Using only images of cells, and the localization information of the fluorescent protein assayed, these methods are able to accurately assign localization for new cells, and to automatically learn which image features are relevant.
- 1393–1404 **Genotype Calling from Population-Genomic Sequencing Data**
Takahiro Maruki and Michael Lynch
- To take maximum advantage of rapidly improving sequencing technologies, we have developed methods for calling genotypes from population-genomic sequencing data of various organisms. Specifically, we provide a Bayesian method for calling diploid genotypes useful for improving the accuracy of genotype calls from low-coverage data. We also provide another type of method suitable for high-coverage data, which is highly efficient and relaxes the assumption of biallelic polymorphisms made in many existing methods. We extend the latter method to triploid and tetraploid data, showing that polyploid data require much higher coverage for accurate genotype calls than diploid data.
- 1405–1416 **A New Reference Genome Assembly for the Microcrustacean *Daphnia pulex***
Zhiqiang Ye, Sen Xu, Ken Spitze, Jana Asselman, Xiaoqian Jiang, Matthew S. Ackerman, Jacqueline Lopez, Brent Harker, R. Taylor Raborn, W. Kelley Thomas, Jordan Ramsdell, Michael E. Pfrender, and Michael Lynch
- We generate a de novo assembly of the genome of a natural *Daphnia pulex* isolate (PA42) with up to 650× total coverage and then compare it with the first sequenced genome of this species (TCO), which was derived from an isolate from a population with >90% reduction in nucleotide diversity. We identified ~7,000 excess genes in TCO resulting from overly aggressive gene annotation. In addition, we observe more proliferation of Transposable elements and a higher frequency of gained introns in the TCO genome that is consistent with the view that reduced effective population size can facilitate the accumulation of excess DNA.
- 1417–1427 **Dynamic Notch Signaling Specifies Each Cell Fate in *Drosophila* Spermathecal Lineage**
Wei Shen and Jianjun Sun
- Genetic manipulation demonstrated surprisingly different ligand mechanism for Notch activation and the essential role of Notch signaling in cell-fate specification in spermathecal lineage. In addition, we illustrated the distinct morphology and function of each cell in the class-III secretory unit in spermathecae.

- 1429–1437 **Reliable CRISPR/Cas9 Genome Engineering in *Caenorhabditis elegans* Using a Single Efficient sgRNA and an Easily Recognizable Phenotype**
Sonia El Mouridi, Claire Lecroisey, Philippe Tardy, Marine Mercier, Alice Leclercq-Blondel, Nora Zariohi, and Thomas Boulin
 CRISPR/Cas9 genome engineering has revolutionized gene editing in the nematode *Caenorhabditis elegans*. El Mouridi *et al.* describe an optimized strategy that facilitates and increases the reliability of CRISPR experiments in the worm. They further use this approach to demonstrate how the detection of fusion proteins expressed at physiological levels can be greatly improved using a novel red fluorescent protein. Indeed, the *C. elegans*-optimized wrmScarlet is 8 times brighter than TagRFP-T when fused to a muscle-expressed potassium channel.
- 1439–1450 **The Interaction of Genetic Background and Mutational Effects in Regulation of Mouse Craniofacial Shape**
Christopher J. Percival, Pauline Marangoni, Vagan Tapaltsyan, Ophir Klein, and Benedikt Hallgrímsson
 We investigated the strength of epistatic contributions to craniofacial shape, a complex trait of interest to researchers studying human birth defects and vertebrate evolution. As a mammalian model of epistatic effects, we measured how three common inbred mouse backgrounds modulate the craniofacial effects of three related Sprouty null mutations. Inbred background modified the strength and direction of craniofacial dysmorphology associated with these mutations. In fact, the interaction of mutation identity and inbred background accounted for as much skull shape variation as mutation alone, suggesting that epistatic interactions are critical in determining complex trait morphology in mammals.
- 1451–1461 **Aging Effects of *Caenorhabditis elegans* Ryanodine Receptor Variants Corresponding to Human Myopathic Mutations**
Katie Nicoll Baines, Célia Ferreira, Philip M. Hopkins, Marie-Anne Shaw, and Ian A. Hope
 Mutations in the human ryanodine receptor gene RYR1 lead to muscle disorders such as malignant hyperthermia. Equivalent changes in the corresponding *Caenorhabditis elegans* gene led to alterations in movement after challenge with the pharmacological agents caffeine and halothane, reflecting observations seen for humans. The single amino acid changes generated by these mutations also led to reduced lifespan and accelerated muscle aging, supporting the hypothesis that failure to maintain calcium ion balance in muscle cells appropriately contributes to reduced mobility in human old age. The change in response to caffeine also depended upon a neural component.
- 1463–1473 ***Schizosaccharomyces pombe* MutS α and MutL α Maintain Stability of Tetra-Nucleotide Repeats and Msh3 of Hepta-Nucleotide Repeats**
Desirée Villahermosa, Olaf Christensen, Karen Knapp, and Oliver Fleck
 Defective mismatch repair in humans is associated with colon cancer and instability of microsatellites, DNA sequences with one or several nucleotides repeated. We analyzed microsatellite stability in *Schizosaccharomyces pombe*. Mismatch repair mutants, defective in MutS α or MutL α caused instability of tetra-nucleotide repeats, but not of penta- or hepta-nucleotide repeats. Thus, MutS α /MutL α -dependent mismatch repair in *S. pombe* is limited to up to four unpaired nucleotides. Loss of Msh3 or Exo1 caused instability of hepta-nucleotide repeats. Inactivation of rad51 or exo1 suppressed increased mutation rates in msh3 mutants. Thus, Msh3 prevents Rad51 and Exo1 driven errors in such repeats.
- 1475–1485 **Extensive Copy Number Variation in Fermentation-Related Genes Among *Saccharomyces cerevisiae* Wine Strains**
Jacob Steenwyk and Antonis Rokas
 Wine strains of *Saccharomyces cerevisiae* harbor low genetic diversity in the form of single nucleotide polymorphisms (SNPs); whether that's also the case for other types of genetic variation remains an open question. We examined if 132 wine yeast strains showing low SNP diversity harbored similarly low levels of copy number (CN) variation. We found that ~4% of the genome of each strain was CN diverse, and included multiple genes involved in fermentation-related processes (e.g., copper resistance, flocculation, and glucose metabolism). We conclude that CN variation significantly contributes to the genome diversity of wine yeasts and likely influences fermentation-related processes.

- 1487–1497 **Paternal Induction of Hybrid Dysgenesis in *Drosophila melanogaster* Is Weakly Correlated with Both P-Element and *hobo* Element Dosage**
Satyam P. Srivastav and Erin S. Kelleher
- Transposable elements (TEs) are virtually ubiquitous components of genomes, yet they often impose significant fitness consequences on their hosts. These fitness costs are often assumed to be dosage-dependent, with stronger effects occurring in the presence of higher TE copy numbers. We test this assumption in *Drosophila melanogaster* by considering the relationship between the copy number of P-element and hobo element DNA transposons and the incidence of hybrid dysgenesis, a sterility syndrome associated with transposon activity in the germline. We find that heritable variation in hybrid dysgenesis is abundant, but that little of this variation can be attributed to transposon dosage.
- 1499–1510 **Evaluating Methods of Updating Training Data in Long-Term Genomewide Selection**
Jeffrey L. Neyhart, Tyler Tiede, Aaron J. Lorenz, and Kevin P. Smith
- A breeding program implementing genomewide selection must update the data in the training population to sustain accurate predictions. In a simulation experiment, we tested six different methods of updating training data over cycles of selection. We found that not updating the training population was highly unfavorable for long-term prediction accuracy and response to selection. Of the tested methods, the simplest one resulted in high prediction accuracy and high response to selection. This suggests that breeders using genomewide selection can update their training population by including routinely-generated data.
- 1511–1524 **Modulating Crossover Frequency and Interference for Obligate Crossovers in *Saccharomyces cerevisiae* Meiosis**
Parijat Chakraborty, Ajith V. Pankajam, Gen Lin, Abhishek Dutta, Krishnaprasad G. Nandanam, Manu M. Tekkedil, Akira Shinohara, Lars M. Steinmetz, and Nishant K. Thazath
- The baker's yeast makes ~90 crossovers per meiosis to ensure an obligate crossover per homolog pair that promotes disjunction. To test the relevance of crossover frequency in ensuring obligate crossovers we used mutations in the crossover control genes *MLH3* and *PCH2* that affect crossover frequency and placement. We show that the obligate crossover is strongly compromised in *mlh3Δ* (64 crossovers per meiosis), and *mlh3Δ pch2Δ* (100 crossovers per meiosis) but weakly affected in *pch2Δ* (137 crossovers per meiosis). These results suggest a random distribution of wild-type crossover frequency (*mlh3Δ pch2Δ*), does not ensure the obligate crossover event.
- 1525–1537 **Genomic Comparison of Indigenous African and Northern European Chickens Reveals Putative Mechanisms of Stress Tolerance Related to Environmental Selection Pressure**
Damarius S. Fleming, Steffen Weigend, Henner Simianer, Annett Weigend, Max Rothschild, Carl Schmidt, Chris Ashwell, Mike Persia, James Reecy, and Susan J. Lamont
- The study examined genomic variation of African and European chicken populations to elucidate how selective pressure has influenced adaptations at the genomic level that allow for survival in contrasted environments. The study results indicated that environmental factors, such as climate, may have influenced selection for some of the genomic divergence in developmental and metabolic processes between these distinct populations. This understanding of the genomic response to non-optimal environments can lend new insight into the genes under possible selection for tolerance and survival and, therefore, provide a foundation for determining the influence of environment on genetic contributions to adaptive traits.
- 1539–1549 **TheCellMap.org: A Web-Accessible Database for Visualizing and Mining the Global Yeast Genetic Interaction Network**
Matej Usaj, Yizhao Tan, Wen Wang, Benjamin VanderSluis, Albert Zou, Chad L. Myers, Michael Costanzo, Brenda Andrews, and Charles Boone
- This study describes, TheCellMap.org, a web-accessible database and visualization tool, designed to facilitate access and navigation of the global yeast genetic interaction network.

- 1551–1561 **Genotyping-by-Sequencing Facilitates a High-Density Consensus Linkage Map for *Aegilops umbellulata*, a Wild Relative of Cultivated Wheat**
Erena A. Edae, Pablo D. Olivera, Yue Jin, and Matthew N. Rouse
 A high density consensus map allows precisely identifying QTL position for traits of interest. It can also be used to anchor and order scaffold assemblies to develop genome reference. *Aegilops umbellulata* has been used as source of abiotic and biotic resistance genes in wheat improvement programs. However, both high density consensus map and reference genome sequences have not been reported yet. In the current work, we developed high density linkage map for *Ae. umbellulata* that can facilitates not only QTL mapping but also anchoring and ordering genome assemblies for this species.
- 1563–1568 **From Pine Cones to Read Clouds: Rescaffolding the Megagenome of Sugar Pine (*Pinus lambertiana*)**
Marc W. Crepeau, Charles H. Langley, and Kristian A. Stevens
 Conifers represent the largest reference genome projects executed to date. The largest of these, is that of the conifer *Pinus lambertiana* (sugar pine), with a genome size of 31 billion bp. In this paper, we report on the molecular and computational protocols for scaffolding the *P. lambertiana* genome using the library technology from 10X Genomics. At 247 thousand base pairs, the NG50 of existing reference sequence is the highest scaffold contiguity among the currently published conifer assemblies; this new assembly's NG50 is 1.94 million base pairs, an eightfold increase.
- 1569–1575 **asymptoticMK: A Web-Based Tool for the Asymptotic McDonald–Kreitman Test**
Benjamin C. Haller and Philipp W. Messer
 The McDonald–Kreitman test is a widely used method for quantifying the role of positive selection in molecular evolution, yet its estimates can be severely biased by slightly deleterious mutations. An asymptotic version of this test was recently introduced that addresses this problem by evaluating polymorphism levels for different mutation frequencies separately, and then extrapolating a function fitted to that data. Here we present asymptoticMK, a web-based implementation of this asymptotic McDonald–Kreitman test, which is free to use, open-source, and available at <http://benhaller.com/messerlab/asymptoticMK.html>.
- 1577–1585 **Distinct Patterns of Gene Gain and Loss: Diverse Evolutionary Modes of NBS-Encoding Genes in Three Solanaceae Crop Species**
Lan-Hua Qian, Guang-Can Zhou, Xiao-Qin Sun, Zhao Lei, Yan-Mei Zhang, Jia-Yu Xue, and Yue-Yu Hang
 447, 255 and 306 NBS-encoding genes were identified from the genomes of potato, tomato and pepper, respectively. Three monophyletic clades (TNLs, CNLs and RNLs) were formed by these genes, which derived from 150 CNL, 22 TNL and four RNL ancestral genes and exhibit diverse evolutionary patterns. Potato shows a “consistent expansion” pattern, tomato exhibits a pattern of “first expansion and then contraction”, and pepper presents a “shrinking” pattern. The earlier expansion of CNLs in the common ancestor led to the dominance of this subclass in gene numbers. Species-specific tandem duplications contributed the most to NBS-encoding gene expansions in Solanaceae.
- 1587–1594 **Identification of QTLs for 14 Agronomically Important Traits in *Setaria italica* Based on SNPs Generated from High-Throughput Sequencing**
Kai Zhang, Guangyu Fan, Xinxin Zhang, Fang Zhao, Wei Wei, Guohua Du, Xiaolei Feng, Xiaoming Wang, Feng Wang, Guoliang Song, Hongfeng Zou, Xiaolei Zhang, Shuangdong Li, Xuemei Ni, Gengyun Zhang, and Zhihai Zhao
 Foxtail millet (*Setaria italica*) is an important crop possessing C4 photosynthesis capability. In the present study, we resequenced a foxtail millet population of 439 recombinant inbred lines (RILs), developed high resolution bin map and high density SNP markers, of which could provide an effective approach for gene identification. A total of 59 QTLs for 14 agronomic traits in plants grown under long day and short day photoperiods were identified. The phenotypic variation explained ranged from 4.9% to 43.94%. The newly identified QTLs will provide a platform for sequence-based research on the *S. italica* genome, and for molecular assisted breeding.

- 1595–1606 **A Bayesian Poisson-lognormal Model for Count Data for Multiple-Trait Multiple-Environment Genomic-Enabled Prediction**
Osva A. Montesinos-López, Abelardo Montesinos-López, José Crossa, Fernando H. Toledo, José C. Montesinos-López, Pawan Singh, Philomin Juliana, and Josafhat Salinas-Ruiz
- When a plant scientist wishes to make genomic-enabled predictions of multiple traits measured in multiple individuals in multiple environments, the most common strategy for performing the analysis is to use a single trait at a time taking into account genotype \times environment interaction (G \times E), because there is a lack of comprehensive models that simultaneously take into account the correlated traits and G \times E. In this study we propose a Bayesian multiple-trait and multiple-environment model for count data. Results show that the proposed multi-trait, multi-environment model is an attractive alternative for modeling multiple count traits measured in multiple environments.
- 1607–1615 **How Changes in Anti-SD Sequences Would Affect SD Sequences in *Escherichia coli* and *Bacillus subtilis***
Akram Abolbaghaei, Jordan R. Silke, and Xuhua Xia
- Translation initiation in prokaryotes involves initiation signals such as start codon and Shine-Dalgarno (SD) sequence being decoded by the initiation tRNA and anti-SD sequence in the 3' end of small subunit rRNA (3'TAIL). Even a minor difference in 3' TAIL between *Escherichia coli* and *Bacillus subtilis* (three extra nucleotides in *B. subtilis* leads to different SD/aSD co-adaptation and differential SD usage between the two species. The differences in SD/aSD contribute to differences in host specificity in bacteriophage. Highly expressed genes (HEGs) use longer SDs than lowly expressed genes (LEGs), more so in *Bacillus subtilis* than in *Escherichia coli*. HEGs in both species have more SDs matching to the core motif ACCUCCUU of 3'TAIL than The three extra nucleotides in *B. subtilis* 3'TAIL is used as part of aSD mainly by LEGs.