BioNet

BIOINFORMATICS
NETWORK
ALBERTA

The 3rd Annual BioNet Conference

May 31st to June 2nd , Lister Conference Center
University of Alberta

The Western Canada Bioinformatics and Omics conference

Talks. Networking. Poster Sessions.

Conference Program

Welcome to the 3rd Annual BioNet Conference

Dear members of the BioNet community,

On behalf of the organizing committee, we would like to welcome you to the 3rd Annual BioNet Alberta Bioinformatics and Omics Conference. BioNet Alberta is a diverse network-based approach to building provincial capacity in the area of Bioinformatics and Computational Biology. We seek to achieve these objectives by fostering an environment of collaboration and cooperation among researchers across a wide variety of disciplines. This conference represents an important physical step in bringing together scientists from institutions across Alberta and the Western provinces to identify and promote services, platforms and potential partnerships that may be mutually beneficial. We hope you enjoy this year's conference!

Territorial acknowledgement:

BioNet Alberta respectfully acknowledges that this conference is located on Treaty 6 territory, a traditional gathering place for diverse Indigenous peoples including the Cree, Blackfoot, Métis, Nakota Sioux, Iroquois, Dene, Ojibway/ Saulteaux/Anishinaabe, Inuit, and many others whose histories, languages, and cultures continue to influence our vibrant community.

3rd Annual BioNet Conference Organizing Committee:

Eric Merzetti Angeliki Pantazi Athanasios Zovoilis

3rd Annual BioNet Conference Scientific Organizing Committee:

Jason de Koning Paul Gordon Emily Herman Athan Zovoilis

3rd Annual BioNet Conference Poster judges:

Emily Herman Vanessa Carias

3rd Annual BioNet Conference Student Organizing Committee:

Luke Saville Riya Roy Travis Haight Elly Wu Jemaneh Habtewold

Liam Mitchell Alexey Sizykh Sani-e-Zehra Zaidi Matthew Steward-Edwards

3rd Annual BioNet Conference Industrial Sponsors:

illumina®

PacBio

3rd Annual BioNet Conference Sponsors:

BioNet Prairie

BIONET BIOINFORMATICS NETWORK ALBERTA

BioNet would like to thank our network sponsors for their continuous support:

BioNet Founders and Sponsors

UNIVERSITY OF CALGARY

University of Lethbridge

UNIVERSITY OF ALBERTA

GenomeAlberta

GenomeCanada

ALBERTA INNOVATES

UNIVERSITY OF CALGARY CUMMING SCHOOL OF MEDICINE Centre for Health Informatics

Southern Alberta Genome Sciences Centre

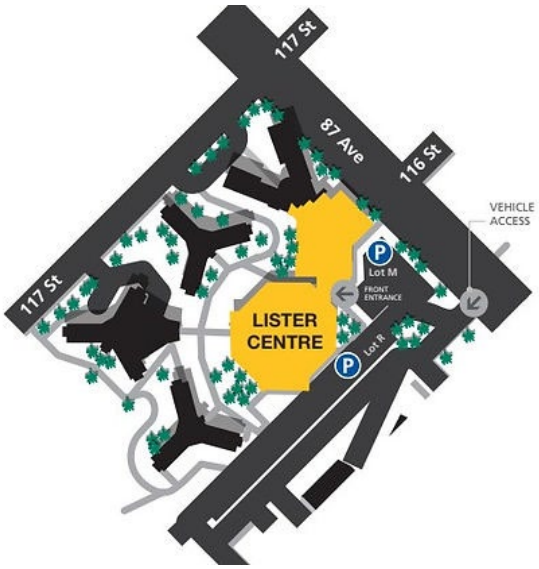
Alberta Government

ALBERTA PRECISION LABORATORIES Leaders in Laboratory Medicine

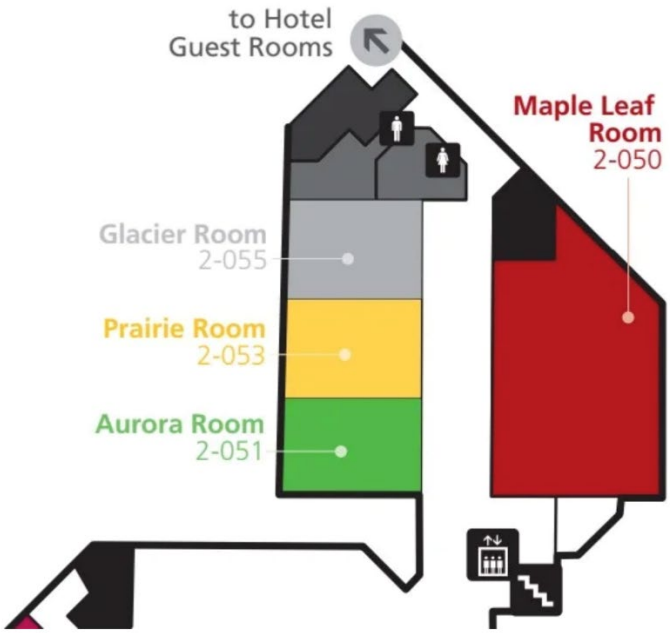
Agriculture and Agri-Food Canada

BioNet Map and Location Information:

The 3rd Annual BioNet Conference will be held on the North Campus of the University of Alberta at the Lister Conference Center (LH). The Lister Centre is located at 87 Avenue and 116 Street. There is dedicated parking (Lot M) in front of the conference center and a 5 minute walk to the Health Sciences LRT Station.

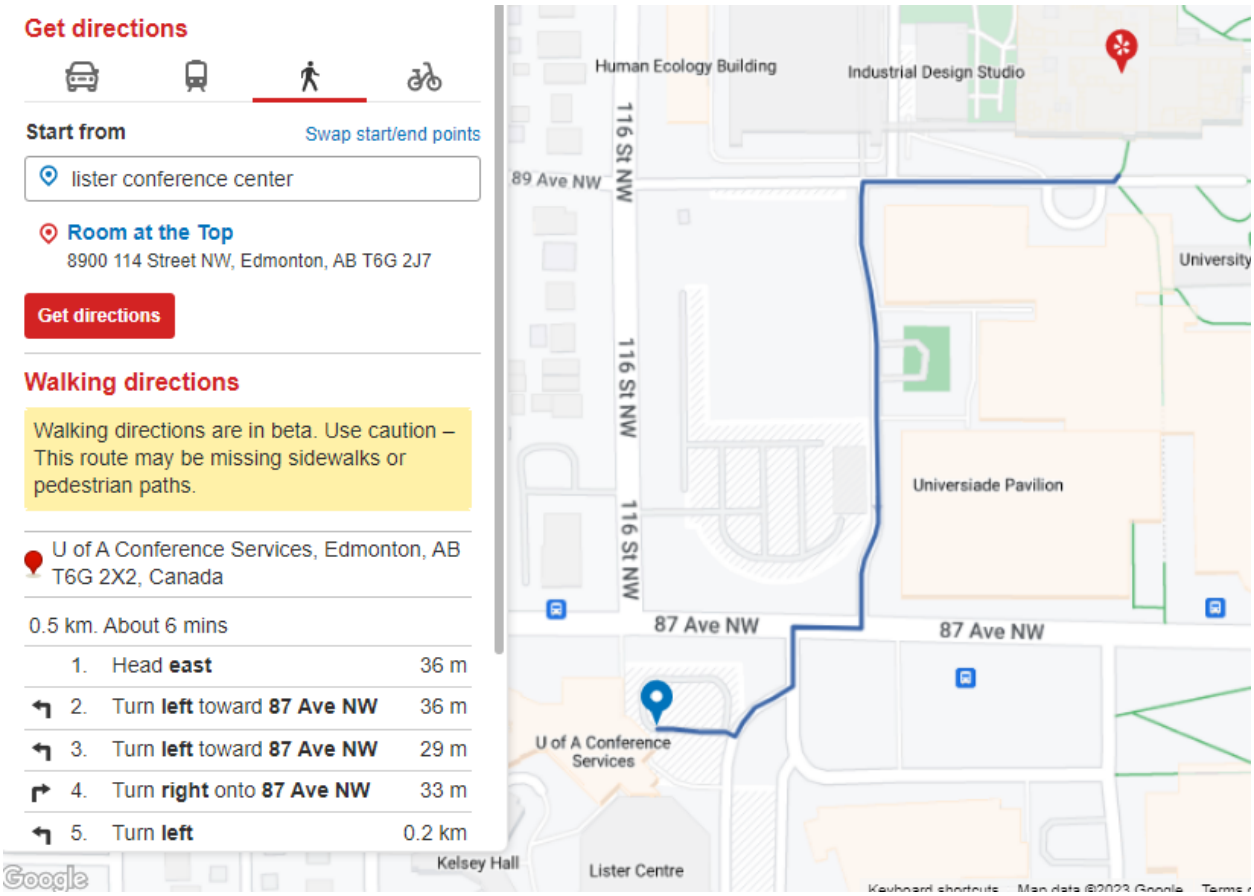


The main conference will be held in the Maple Leaf Room (2-050) which is on the 2nd floor of the Lister Center. The Aurora (2-051) and Prairie (2-053) rooms will be used for breakout sessions and poster presentations. Poster presentations will be broken up by room with corresponding numbers indicated within the conference program.



The Room at the Top:

The room at the top is the location for our closing banquet and located on the 7th floor of the Students' Union Building.



The 3rd Annual BioNet Conference –
Program at a glance

Wednesday, May 31st

9:00am-12:00pm	Alberta Agricultural Omics Get Together	Maple Leaf Room
12:00am-1:00pm	Registration	Outside Maple Leaf Room
1:00pm-1:05pm	Welcome and Opening Remarks	Maple Leaf Room
1:05pm-1:10pm	BioNet Introduction	Maple Leaf Room
1:10pm-1:15pm	Conference Kickoff	Maple Leaf Room
1:15pm-1:45pm	Bioinformatics White Paper Draft Presentation	Maple Leaf Room
1:50-2:15pm	Plenary: Dr. Russ Greiner	Maple Leaf Room
2:20pm-3:00pm	Session 1: Agricultural Omics	Maple Leaf Room
3:00pm-3:30pm	Refreshment Break	Maple Leaf Room
3:30pm-4:55pm	Session 2: Agricultural Omics 2	Maple Leaf Room
5:00pm-6:00pm	Keynote: Dr. Michael Hoffman	Maple Leaf Room
6:00pm-7:00pm	Dinner	Maple Leaf Room

Thursday, June 1st

8:00am-9:00am	Breakfast	Maple Leaf Room
9:00am-9:25am	Plenary: Dr. Steven Jones	Maple Leaf Room
9:30am-11:00am	Session 3: Health Omics	Maple Leaf Room
11:00am-11:30am	Refreshment Break	Maple Leaf Room
11:30am-12:50pm	Session 4: Advanced Bioinformatic Approaches	Maple Leaf Room
1:00pm-2:00pm	Lunch	Maple Leaf Room

2:00-3:00pm	Even Numbered Posters Present	Aurora and Prairie Rooms
3:00pm-4:00pm	Session 5: Health Omics 2	Maple Leaf Room
4:00pm-4:30pm	Refreshment Break	Maple Leaf Room
4:30pm-4:55pm	Plenary: Dr. David Wishart	Maple Leaf Room
5:00pm-6:00pm	Session 6: Bioinformatics Workflows	Maple Leaf Room
6:00pm-7:00pm	Dinner	Maple Leaf Room

Friday, June 2nd

8:00am-9:00am	Breakfast	Maple Leaf Room
9:00am-9:25am	Plenary: Dr. Gwendolyn Blue	Maple Leaf Room
9:30am-10:55am	Session 7: Agricultural Omics 3	Maple Leaf Room
11:00am-11:50am	Extended Networking Break	Maple Leaf Room
11:50am-1:30pm	Session 8: Microbial Omics	Maple Leaf Room
1:30pm-2:30pm	Lunch	Maple Leaf Room
2:30-3:30pm	Odd Numbered Posters Present	Aurora and Prairie Rooms
3:30pm-4:30pm	PI / Student Meetings	Maple Leaf Room / Prairie Room
4:30pm-5:00pm	Refreshment Break / Group Photo	Maple Leaf Room
5:00pm-6:00pm	Keynote: Dr. Alice Berger	Maple Leaf Room
6:00pm-7:00pm	Dinner / Awards / Final Comments	Maple Leaf Room
7:00pm-10:00pm	Banquet	Room at the Top

3rd Annual BioNet Conference – Full Program

Wednesday, May 31st

Check in and badge pickup starting at 12pm in Hallway outside Maple Leaf Room

Opening Ceremonies

1:00 – 1:05pm	**Dr. Athanasios Zovoilis** Academic Lead, BioNet Alberta	
1:05 – 1:10pm	**Dr. André McDonald, Associate Vice-President (Strategic Research Initiatives and Performance)** University of Alberta	
1:10 – 1:15pm	**Opening Remarks** Genome Alberta	
1:15 – 1:45pm	**Bioinformatics White Paper Draft Presentation** Dr. Athan Zovoilis and Dr. Eric Merzetti	

Opening Plenary

Chair: Dr. Eric Merzetti

1:50 – 2:15 pm	**Dr. Russ Greiner, The University of Alberta / AMII** Intro to machine learning for medical researchers / practicians	**2**

Session 1: Agricultural Omics

Chair: Dr. Tim McAllister

2:20 – 2:40 pm	**Dr. Devin Holman, AAFC Lacombe** Early-life antibiotic treatment and the piglet gut microbiome and resistome	**4**
2:45 – 3:00 pm	**Lance Lansing, AAFC Lethbridge** Taxonomic classification and analysis workflow for microbiome sequence data using Snakemake, R, and R markdown	**32**
3:00 – 3:30 pm	**Refreshment Break**	

3:00 – 3:30 pm	**Refreshment Break**	

Session 2: Agricultural Omics 2

Chair: Rodrigo-Ortega Polo

3:30 – 3:50 pm	**Dr. Tim McAllister, AAFC Lethbridge** Genomic and metagenomic approaches to defining the ecology of AMR from a One Health perspective	**4**
3:55 – 4:15 pm	**Dr. Robert Gruninger, AAFC Lethbridge** Understanding the host –microbiome interactions involved in liver abscesses formation in beef cattle	**5**
4:20 – 4:40 pm	**Dr. Leluo Guan, University of Alberta** Translational rumen microbiome research for strategies to improve cattle production sustainability	**5**
4:40 – 4:55 pm	**Dr. Jason Grant, University of Alberta** Proksee: a web server for the characterization and visualization of bacterial genomes	**68**

Opening Keynote

Chair: Dr. Athan Zovoilis

5:00 – 6:00 pm	**Dr. Michael Hoffman, University of Toronto** Identifying transcription factor binding using open chromatin, transcriptome, and methylation data	**1**
6:00 – 7:00 pm	**Supper**	

Thursday, June 1st

8:00 – 9:00 am **Breakfast**

Plenary Talk

Chair: Dr. Pinaki Bose

9:00 – 9:25 am	**Dr. Steven Jones, Professor, FRSC, FCAHS** Studying Genomes when the Epigenome comes along for the ride	**2**

Session 3: Health Omics

Chair: Dr. Ly Vu

9:30 – 9:50 am	**Dr. Pinaki Bose, University of Calgary** Living on the Edge: How Distinct Cancer Cell States Drive Tumour Aggressiveness	**6**
9:55 – 10:15 am	**Dr. Ly Vu, University of British Columbia** Toward targeting post-transcriptional gene expression control in leukemia	**6**
10:20 – 10:40 am	**Dr. Gane Ka-Shu Wong, University of Alberta** Drug Target Discovery In The Era Of Digital Health Databases	**7**
10:45 – 11:00 am	**Dr. Mahmoud M. Mustafa, University of Calgary** Cooperative Interactions of Glucocorticoid Receptor and NF-κB at Overlapping Genomic Loci: Insights from ChIP-seq Analysis	**43**
11:00 – 11:30 pm	**Refreshment Break**	

Session 4: Advanced Bioinformatic Approaches

Chair: Dr. Paul Gordon

11:30 – 11:50 am	**Dr. Jason de Koning, University of Calgary** Why are Human-Disease Associated Variants So Common in the Genomes of Non-Human Vertebrates?	**8**
11:55 – 12:15 pm	**Dr. Quan Long, University of Calgary** CATE: A fast and scalable CUDA-based tool characterizing molecular evolution	**8**
12:20 – 12:35 pm	**Dr. Tatiana Maroilley, University of Calgary** A CURE for Introduction to Bioinformatics	**39**
12:40 – 12:55 pm	**David Enoma, University of Calgary** A statistical method for image-mediated association studies discovers genes and pathways associated with four brain disorders	**18**

1:00 – 2:00 pm	**Lunch**	
2:00 – 3:00 pm	**Poster Session 1 – Even Numbered Posters Present (Prairie and Aurora Rooms)**	

Session 5: Health Omics 2

Chair: Dr. Pinaki Bose

3:00 – 3:20 am	**Dr. Matthew Croxen, Alberta Precision Laboratories** Modernizing Alberta's Public Health Laboratory Using Genomics	**8**
3:25 – 3:40 pm	**Anureet Tiwana, University of Alberta** Identifying Key Genes in Preeclampsia using Machine Learning	**54**
3:45 – 4:00 pm	**Travis Haight, University of Lethbridge** Shining light on circRNA dysregulation in neurogenerative disease	**23**
4:00 – 4:30 pm	**Refreshment Break**	

Plenary Talk

Chair: Jason de Koning

4:30 – 4:55 am	**Dr. David Wishart, University of Alberta** Bioinformatics and Biomarkers	**3**

Session 6: Bioinformatic workflows

Chair: Dr. Jason de Koning

5:00 – 5:20 pm	**Dr. Paul Gordon, University of Calgary** From swab to sequence: towards an on-site, precision antimicrobial treatment workflow for cattle feedlots	**9**
5:25 – 5:40 pm	**Dr. Rui (Ric) Qin, The Metabolomics Innovation Centre** Biomarker discovery workflow for chemical isotope labeling LC-MS metabolomics	**48**
5:45 – 6:00 pm	**K M Tahsin Hassan Rahit, University of Calgary** ModSpy: A Machine Learning model detects Genetic Modifiers from Whole Genome Sequencing data of Model Organisms	**66**
6:00 – 7:00 pm	**Supper**	

Friday, June 2nd

8:00 – 9:00 am **Breakfast**

Plenary Talk

Chair: Dr. Paul Gordon

9:00 – 9:25 am	**Dr. Gwendolyn Blue, University of Calgary** Responsible research and innovation: From principles to practice	**3**

Session 7: Agricultural Omics 3

Chair: Dr. Devin Holman

9:30 – 9:50 am	**Rodrigo Ortega-Polo, AAFC Lethbridge** Lethbridge CARSU and the Bioinformatics Research Support Network: Accelerating the Digital Transformation of the Agriculture and Agri-Food sector at AAFC	**9**
9:55 – 10:15 am	**Dr. Lingling Jin, University of Saskatchewan** Genome evolution and reconstruction of ancestral monoploid karyotypes	**10**
10:20 – 10:35 am	**Greg Robinson, University of Lethbridge** Copper Sulphate Foliar Applications and Soil Amendments Alters Pathogens Found on Cannabis sativa	**69**
10:40 – 10:55 am	**Keyhan Najifian, University of Saskatchewan** Semi-Self-Supervised Learning for Semantic Segmentation in Images with Dense Patterns	**46**
11:00 – 11:50 pm	**Extended Networking Break**	

Session 8: Advanced Bioinformatic Approaches

Chair: Dr. Matthew Croxen

11:50 – 12:10 pm	**Dr. Elisabeth Richardson, Mount Royal University** Comparative genomics of a pan-ciliate dataset reveal paradoxical protein complexes associated with membrane trafficking and peroxisome formation	**11**

12:15 – 12:30 pm	**Dr. Viraj Muthye, University of Calgary** Jumping jellyfish: Myxobolusrasmusseni, a cnidarian parasite of Albertan fish, has a genome that is highly repetitive, full of transposons, and has experienced massive gene loss	**45**
12:35 – 12:50 pm	**Varada Khot, University of Calgary** Decoding genomic adaptations of an alkaliphilic Opitutalesusing gene-tree species-tree reconciliation	**30**
12:55 – 1:10 pm	**Daniel A. Salazar-Alemán, University of Calgary** Toxicometallomics of Escherichia coli grown in the presence of silver, copper, and gallium metal salts: similarities and differences at the transcriptome level	**12**
1:15 – 1:30 pm	**Daniel Yu, University of Alberta** Functional genomic characterization of naturalized Escherichia coli strains adapted to food-and water-associated industrial environments	**61**
1:30 – 2:30 pm	**Lunch**	
2:30 – 3:30 pm	**Poster Session 2 – Odd Numbered Posters Present (Prairie and Aurora Rooms)**	
3:30 – 4:30 pm	**PI (Maple Leaf Room) / Student Meetings (Aurora Room)**	
4:30 – 5:00 pm	**Refreshment Break / Group Photo (Maple Leaf Room)**	

<u>Closing Keynote</u>

Chair: Dr. Athan Zovoilis

5:00 – 6:00 pm	**Dr. Alice Berger** CRISPR-based discovery of cancer drug targets to enable precision medicine	**1**
6:00 – 7:00 pm	**Supper / Awards**	
7:00 – 10:00 pm	**Conference Banquet (Room at the Top)**	

3rd Annual BioNet Conference – Poster Presentations

The 3rd Annual BioNet Conference
Keynote and Plenary Speakers

Dr. Michael Hoffman, PhD

Identifying transcription factor binding using open chromatin, transcriptome, and methylation data

Michael Hoffman creates predictive computational models to understand interactions between genome, epigenome, and phenotype in human cancers. His influential machine learning approaches have reshaped researchers' analysis of gene regulation. These approaches include the genome annotation method Segway, which enables simple interpretation of multivariate genomic data. He is a Senior Scientist in and Chair of the Computational Biology and Medicine Program, Princess Margaret Cancer Centre and Associate Professor in the Departments of Medical Biophysics and Computer Science, University of Toronto. He was named a CIHR New Investigator and has received several awards for his academic work, including the NIH K99/R00 Pathway to Independence Award, and the Ontario Early Researcher Award.

Dr. Alice Berger, PhD

CRISPR-based discovery of cancer drug targets to enable precision medicine

Alice Berger, Ph.D., is an Associate Professor and the Innovators Network Endowed Chair at Fred Hutchinson Cancer Center in Seattle, WA. The Berger laboratory is focused on improving the outcomes of lung cancer patients through discovery of the genomic determinants of lung cancer initiation and therapeutic response.

Dr. Berger earned a B.S. in Chemistry at the University of Virginia and a Ph.D. in Biochemistry and Molecular Biology from Cornell University's Weill Graduate School of Medical Sciences. She then trained as a postdoctoral fellow in Matthew Meyerson's lab at the Broad Institute of MIT and Harvard and Dana-Farber Cancer Institute. In her work with The Cancer Genome Atlas, she discovered and described cancer-associated mutations in lung cancer such as RIT1 and MET exon 14 skipping. She is a co-developer of eVIP, a technology for rapid assessment of the functional impact of genetic variants. Dr. Berger is the recipient of the Devereaux Outstanding Young Investigator award by the Prevent Cancer Foundation for her work on the genomics of lung cancer in women and an NCI MERIT award on her work on RIT1 driven lung cancer.

Dr. Russ Greiner, CIFAR AI Chair

Intro to machine learning for medical researchers / practicians

Russ Greiner focuses on developing and improving applications of machine learning in medicine, providing solutions for specific real-world problems across a range of clinical considerations. He works closely with clinicians and researchers in medicine (in psychiatry, oncology, cardiovascular, diabetes, and other areas), metabolomics and other disciplines to develop data-driven tools that assist practitioners with screening, diagnosis, prognosis, and treatment planning in physical and mental health. Within the field of computational psychiatry, Russ uses machine learning on fMRI (functional magnetic resonance imaging) and other clinical data to develop new ways of diagnosing schizophrenia and for assessing the severity of a range of symptoms. These techniques can also be used across a range of psychiatric disorders including attention deficit hyperactivity disorder and depression. In the area of precision medicine, Russ works with colleagues in healthcare to develop methods for recommending patient-specific plans for the treatment of diseases such as cancer or diabetes and for predicting individual health outcomes. Russ is also interested in building better algorithms that learn from experience, working to produce more robust and effective machine learning systems.

Dr. Steven Jones, PhD, FRS(C), FCAHS, Director, Head, Bioinformatics, Distinguished Scientist, BC Cancer Research Institute

Title: TBD

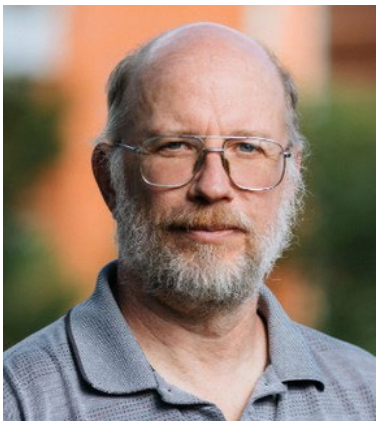
Dr. Jones' research program is firmly entrenched in genome science to better understand the complete mutational landscape of cancers. His primary aim is to help uncover the diversity of genetic and genomic events that accrue to give rise to cancers, and which also encourage their evolution and maintain their progression. His laboratory extensively analyzes Next Generation genome and transcriptome data to achieve these goals. Dr. Jones has developed several novel computational approaches and methodologies to this end and has provided numerous insights into cancer dynamics, potential biomarkers, and therapeutic targets. A significant part of Dr. Jones research program relates to developing more precise cancer treatments by exploiting an individual's specific cancer genome profile. His research has identified numerous epigenetic targets that have the potential to be modulated in such a way as to reverse the effects of mutations within a cancer genome. Using computational approaches, his research team has identified and refined compounds that modify epigenetic programs in cancer. His laboratory also acts as a data analysis center for the Canadian Epigenetics, Environment and Health Research Consortium (CEEHRC).

Dr. Gwendolyn Blue, PhD

Responsible research and innovation: From principles to practice

Recent years have seen an increasing interest in voluntary ethical principles to guide the development of emerging technologies such as artificial intelligence and genomics. The proliferation of principles, including responsible research and innovation (RRI), data ethics, and EDI (equity, diversity, and inclusion) are, in part, a response to high-profile scandals that threaten to undermine public trust and support for research and innovation. Critics argue that the proliferation of voluntary principles is a form of 'ethics washing', where lip service is paid to virtues while status quo practices are retained. Outlining a brief history of calls for responsible research in biotechnology, this presentation calls for a re-envisioning of ethical principles, not as abstract moral virtues, but as practical tools to help guide difficult decisions in the face of conflicting values, needs, and demands. Expertise, experience, and capacity building in negotiating trade-offs, acknowledging uncertainty, and fostering public debate are essential to expedite and scale principle-focused research and innovation: virtues alone are not enough.

Gwendolyn Blue is a Professor in the Department of Geography at the University of Calgary. Trained in microbiology and cultural studies, her research examines the social contexts and power dynamics that inform public debates about issues that involve science and technology, including climate change and genomics. Current research examines public assessments of emerging technologies, with a focus on calls for responsible research and innovation for gene editing of nonhuman organisms.



Dr. David Wishart, PhD

Bioinformatics and Biomarkers

Dr. David Wishart (PhD Yale, 1991) is a Distinguished University Professor in the Departments of Biological Sciences and Computing Science at the University of Alberta. He also holds adjunct appointments with the Faculty of Pharmaceutical Sciences and with the Department of Pathology and Laboratory Medicine. He has been with the University of Alberta since 1995. Dr. Wishart's research interests are very wide ranging, covering bioinformatics, precision medicine, metabolomics, analytical chemistry, drug chemistry, natural product chemistry, molecular biology, protein chemistry and neuroscience. Dr. Wishart has led the "Human Metabolome Project" (HMP), a multi-university, multi-investigator project that is cataloguing all the known chemicals in human tissues and biofluids. This information has been archived on a freely accessible web-resource called the Human Metabolome Database (HMDB). Dr. Wishart has also been using machine learning and artificial intelligence to help create other useful databases, such as DrugBank, MarkerDB, NP-MRD, FooDB and ContaminantDB and software tools (such as Proteus2, MetaboAnalyst, CFM-ID and BioTransformer) to help with the characterization and identification of proteins, metabolites, drugs, and biomarkers. Much of this work is being moved into clinical practice and clinical tests to help move the field of precision medicine

further ahead. Over the course of his career Dr. Wishart has published more than 500 research papers in high profile journals on a wide variety of subject areas. These papers have been cited >120,000 times.

The 3rd Annual BioNet Conference
Invited Speakers

Dr. Devin Holman, Agriculture and Agri-Food Canada

Early-life antibiotic treatment and the piglet gut microbiome and resistome

Dr. Devin Holman is currently a Research Scientist in Livestock Microbiology at Agriculture and Agri-Food Canada based at the Lacombe Research and Development Centre (RDC) in Lacombe, AB. He has a M.Sc. degree in Food Microbiology from Dalhousie University in Halifax, NS, and a Ph.D. in animal science and microbiology from McGill University in Montreal, QC. Dr. Holman was a postdoctoral fellow with AAFC at the Lethbridge RDC focusing on the nasal microbiome and its association with bovine respiratory disease in beef cattle. He also did a postdoctoral fellowship with the USDA-Agricultural Research Service at the National Animal Disease Center in Ames, IA, where his research involved challenging pigs with multidrug-resistant *Salmonella Typhimurium* and investigating the effect that high-dose antimicrobial administration had on *S. Typhimurium* shedding and on the pig gut microbiome. Since starting with AAFC in Lacombe in June 2017, Dr. Holman has been involved in several swine projects designed to characterize the transfer of antimicrobial resistance genes among pigs and to determine the effect of various management and antimicrobial treatment strategies on the pig gut microbiome. Overall, his work is largely focused on monitoring antimicrobial resistance in food producing animals, reducing antimicrobial use, and improving animal health and performance through microbiome-driven research.

Dr. Tim McAllister, Agriculture and Agri-Food Canada

Genomic and metagenomic approaches to defining the ecology of AMR from a One Health perspective

Tim McAllister was raised on a mixed cow-calf operation in Inniisfail Alberta. He obtained his M.Sc. in Animal Biochemistry at the University of Alberta and his Ph. D. in microbiology and nutrition from the University of Guelph in1991. He is presently a principal research scientist with Agriculture and Agri-Food in Lethbridge, Alberta, Canada. Tim leads a diverse research team that has been studying antimicrobial resistance in beef cattle production systems since 1997. The team's recent work has focused on studying AMR from a "One Health" perspective using enterococci as AMR indicators in beef production, human sewage and clinical settings. He is also investigating the role of integrative conjugative elements in the transfer of antimicrobial

resistance genes within the bacterial bovine respiratory disease complex. Tim has authored over 850 scientific papers, the recipient of several national and international awards and holds adjunct professorship appointments at several universities in Canada and abroad.

Dr. Robert Gruninger, Agriculture and Agri-Food Canada

Understanding the host –microbiome interactions involved in liver abscesses formation in beef cattle

A microbiome is a community of microbes that inhabits a particular environment. These communities can be complex and consist of many different types of microbes including bacteria, archaea, fungi, protozoa, and viruses. The microbial community that inhabits the gut is highly diverse and plays and essential role in digestion and heath. Ruminants (cows, goats, sheep, deer, etc) have a specialized chamber in their gastrointestinal tract called a rumen that functions in the digestion of ingested plant material. The rumen microbiome is responsible for the efficient digestion of plant carbohydrates and the conversion of these sugars into volatile fatty acids which serve as the animal's energy source. The microbes in the rumen are also the primary source of protein used by ruminants for growth. My research is focused on applying state of the art genomic techniques, classical microbiology, and protein biochemistry to understand how the microbes in gastrointestinal tract of cattle influence the health and efficiency of the animal. Ultimately the goal is to improve the health and efficiency of cattle.

Dr. Leluo Guan, University of Alberta

Translational rumen microbiome research for strategies to improve cattle production sustainability

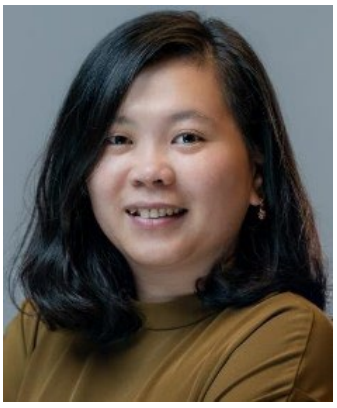
Cattle rumen microbiome plays a critical role in convert fibrous feed stocks to nutrients which directly influence cattle growth and performance. Tremendous research has revealed the composition and function of rumen microbiome under various production and management regime and environment settings and reveal that the difference and/or changes in rumen microbiome can affect many economically important traits such as feed efficiency, milk/meat yield and quality, methane emission, metabolic health and so on. There is an urgent need to combine advanced omics approaches and proper analytic methods as well as validation models to improve microbiome analysis to determine mode of actions of rumen microbiome for each trait. Moreover, omics based translational microbiome solutions are needed to discover practical and adaptable products/tools for future strategies to improve production efficiency, health and welfare, and environment-friendliness in cattle production through microbiome manipulations.

Dr. Pinaki Bose, University of Calgary

Living on the Edge: How Distinct Cancer Cell States Drive Tumour Aggressiveness

Dr. Pinaki Bose completed his Ph.D. in basic cancer research with Dr. Karl Riabowol at the University of Calgary, investigating the role of the ING1 tumour suppressor protein in DNA damage signaling and apoptosis. After completing his Ph.D., Pinaki joined the Ohlson Research Initiative (ORI) as a postdoctoral fellow and trained in the molecular epidemiology of head and neck cancers under the supervision of Drs. Joseph Dort and Nigel Brockton. As part of a second postdoctoral fellowship, Pinaki trained at the BC Cancer Agency (BCCA) Genome Sciences Centre in cancer genomics and bioinformatics under one of the leading bioinformaticians in the world, Dr. Steven Jones. Pinaki was also a member of the personalized oncogenomics (POG) program at the BCCA. The POG initiative administers targeted therapies to recurrent/metastatic cancer patients based on their genomic profiles.

Dr. Bose currently directs the translational research program at the ORI, a multidisciplinary head and neck cancer research initiative working in close collaboration with surgeons, pathologists, bioinformaticians, statisticians and basic researchers. The Bose lab is investigating the biology of head and neck cancers with particular emphasis on mechanisms regulating lymph node metastasis (LNM). LNM is a potent predictor of worse prognosis in oral cancers, one of the most common cancers of the head and neck region. Towards this end, we are developing genome-wide screens in oral cancer cell lines and animal models to identify regulators of LNM.



Dr. Ly Vu, UBC, Terry Fox Laboratory

Toward targeting post-transcriptional gene expression control in leukemia

Dysregulation of normal gene expression programs and cellular identities drive cancer development. While somatic alterations in genetic and epigenetic mechanisms have been studied extensively, how processes that affect post-transcriptional and translational regulation impact tumorigenesis is much less well understood. All transcribed mRNAs are processed with an addition of a poly-A tail, whose length has direct impact on mRNA abundance and translational efficiency. The CCR4-NOT (CNOT) complex is one of two major multi-subunit polyA deadenylation complexes which mediate the shortening of the poly(A) tails. The CNOT complex was found to be recruited to RNA methylation m6A transcripts to mediate mRNA degradation. While deadenylation and the CNOT complex are central to the control of gene expression, their role, and mechanisms of action in malignant cells are poorly understood. Our laboratory uncovered a critical role of CNOT complex in pathogenesis of acute myeloid leukemia (AML), a genetically complex and heterogeneous set of diseases with poor survival and morbidity. Insights into these largely unexplored areas could lead to the development of innovative and more effective treatments for AML patients.

Dr. Gane Ka-Shu Wong, University of Alberta

Drug Target Discovery In The Era Of Digital Health Databases

Professor Gane Ka-Shu Wong was a founding member of two internationally renowned genomics research organizations. The first was at the University of Washington. The other, named Beijing Genomics Institute (BGI or 華大基因), subsequently grew into one of the largest such ventures in the world and hosted a successful IPO in 2017. His h-index is 74, with a total of 37743 citations, and 3840 citations/year in 2021 (Google Scholar, 06 May 2022). He has 38 publications in top-tier journals like Nature Publishing Group and Science. In the 2019/2020/2021 rankings at Clarivate Analytics, he is a Highly Cited Researcher for the Cross Field category with 20 papers since 2012 rated in the top 1% by citations for their respective fields. He is currently a University of Alberta professor, jointly appointed in the Faculty of Medicine and the Faculty of Science.

His research lies at the cutting edge of technologies, high throughput sequencing and computer analyses, that are revolutionizing biology and medicine. Many of these efforts are multidisciplinary international consortiums, which he leads. These include, from the early days of BGI, landmark papers on genomes for indica rice, silkworm, and chicken. More recently, he led the 1KP initiative that sequenced the transcriptomes for a thousand phylodiverse species of green plants (Oct 31 2019 cover of Nature, 2020 Annual Reviews of Plant Biology). He now leads a successor project called 10KP, sequencing complete genomes for 10x as many species. 1KP was a “first-of-kind” project (for all taxa) when launched. It inspired similar projects for other taxa, leading ultimately to the Earth BioGenome Project now sequencing all known eukaryotic species. It also discovered several novel optogenetic proteins (2014 Nature Methods) that are widely used for studies of the mammalian brains in awake behaving animals. Many licenses have been signed, most notably with GenSight Biologics, whose method to restore vision in patients with retinitis pigmentosa was granted fast-track designation by US-FDA regulators in late 2021.

Another component of his research is the introduction of omics technologies into problems of interest for medicine. These include bacterial metagenomics to monitor patient response in fecal microbial transplantation (2017 JAMA, 2016 Frontiers in Microbiology), proviral integration into genomes of patients with autoimmune disease of suspected infectious etiology, single-cell omics to elucidate metastasis progression in breast cancer (2021 Cell Reports Medicine), and novel nanotechnologies for sequencing low abundant DNA/RNA in clinical samples where the target is below PCR limits of detection (2021 Scientific Reports). With the rapid growth in digital health databases, and the breakthroughs in deep learning algorithms, he has been charting a new path in drug target discovery that will increase the likelihoods of success for clinical trials (2021 npj Genomic Medicine) by eliminating our reliance on animal models and cell lines.

Dr. Wong’s diverse interests are a reflection of his life history. Originally born in Hong Kong, he emigrated to Canada at the age of six, and grew up having to navigate two cultures. Scientifically,

he did not start in biology or medicine. His B.A.Sc. at the University of British Columbia included honours in mathematics, physics, electrical engineering, and computer science. His Ph.D. studies at Cornell University were on experimental low temperature physics. While a postdoctoral fellow at the California Institute of Technology, he was recruited to the nascent Human Genome Project and moved to the University of Washington. There, he met the key people who would eventually form BGI, and some of the people he mentored at BGI have since launched their own companies. He admits that luck played a major role in his career successes.

Dr. Jason de Koning, University of Calgary

Why are Human-Disease Associated Variants So Common in the Genomes of Non-Human Vertebrates?

Substantial analytical and computational challenges impede effective exploitation of the massive genomic data sets being generated by next-generation DNA sequencing technologies. To help overcome these impediments, my laboratory is developing novel bioinformatics approaches for the scalable, high-throughput, simultaneous analysis of many genomes using model-based statistical inference. We are particularly interested in bringing modelling advances in molecular evolution and comparative genomics to bear on interpreting personal genomic variation in humans.

Dr. Quan Long, University of Calgary

CATE: A fast and scalable CUDA-based tool characterizing molecular evolution

Quan Long, PhD, was trained in both mathematics and computer science. His current research focuses on phenotype predictions and gene mapping via integration of multiscale omics using statistics and informatics. He is also interested in within-host evolutionary analysis using sequences data that may lead better prediction of disease status such as cancer progression and pathogen antimicrobial resistance. He was a staff R & D engineer analyzing memory leak at IBM Research; then a staff scientist serving for the 1,000 Genomes Project and other evolution-focused projects at the Wellcome Trust Sanger Institute. Afterwards, he assumed the position of a postdoc fellow at the Gregor Mendel Institute, working on methods development as well as real data analysis for NGS-based variants calling, association mapping, and population genetics. Before joining University of Calgary, he was an assistant professor (research track) in Icahn School of Medicine at Mount Sinai, working on phenotype predictions and gene expression networks.

ALBERTA PRECISION LABORATORIES
Leaders in Laboratory Medicine

Dr. Matthew Croxen, Alberta Precision Laboratories

Modernizing Alberta's Public Health Laboratory Using Genomics

Genomics is a powerful tool to help us better understand microorganisms from very different perspectives. Part of my research interests focuses on

genomics of infectious diseases that have severe health outcomes for infants and children. This includes life-threatening infections that can be passed from mom to baby during childbirth or the emergence of childhood diseases that are normally covered by vaccines. The use of genomics allows us to better understand these diseases and provide insight on preventable measures to protect against them.

Dr. Paul Gordon, University of Calgary

From swab to sequence: towards an on-site, precision antimicrobial treatment workflow for cattle feedlots

Paul Gordon, PhD is the Bioinformatics Manager for the Centre for Health Genomics and Informatics in the Cumming School of Medicine at the University of Calgary. Dr. Gordon began working in bioinformatics in 1996, part of the team sequencing the first complete archaeal genome at the National Research Council of Canada. Subsequently he has employed computational methods for the analysis of everything from fungal genomes for biofuel enzyme discovery, to organ transplant rejection detection, cancer immunology, genomic epidemiology, and rational oligonucleotide design, with 58 peer-reviewed papers. He has been working with nanopore data since the first North American case of high pathogenic avian influenza in early 2014, developing and maintaining a number of open-source tools for nanopore data acquisition and analysis since the time.

Rodrigo Ortega-Polo, Agriculture and Agri-Food Canada

Lethbridge CARSU and the Bioinformatics Research Support Network: Accelerating the Digital Transformation of the Agriculture and Agri-Food sector at AAFC

In the rapidly evolving field of bioinformatics, advancements in computational analyses and digital technologies are vital for driving impactful research outcomes for the agricultural sector. As a federal government department, Agriculture and Agri-Food Canada has a Strategic Place for Science which includes three pillars of a renewed strategic direction: mission-driven science, people first strategy and organizational excellence. Within the mission-driven science pillar there are four mission areas, including accelerating the digital transformation of the agriculture and agri-food sector. To foster collaboration, streamline resources, and catalyze digital transformation, the AAFC Lethbridge Computational Analyses Research Support Unit (CARSU) and the Bioinformatics Research Support Network (BRSN) are some of AAFC's initiatives that help drive the digital transformation in agricultural research.

As a node of the national BRSN, Lethbridge CARSU serves as a hub for researchers, providing comprehensive support in computational analyses for a diverse range of biological data. Through cutting-edge tools, methodologies, and expertise, CARSU empowers researchers to extract meaningful insights from large-scale datasets. This talk will delve into CARSU's collaborative framework, emphasizing the importance of multidisciplinary partnerships and knowledge sharing to overcome complex bioinformatics challenges.

Complementing the efforts of Lethbridge CARSU, the AAFC national BRSN acts as a network of bioinformatics research support professionals, fostering a community of practice and knowledge exchange. By facilitating collaboration between researchers and bioinformaticians the BRSN strengthens the bioinformatics research landscape and fosters innovative solutions. Attendees will gain insights into the BRSN's mission, its engagement strategies, and the impact it has on facilitating digital transformation in bioinformatics.

The talk will highlight successful case studies and new projects and initiatives, demonstrating how Lethbridge CARSU and the BRSN have contributed to advancements in diverse research domains within agriculture. This talk will share insights into collaborative networks, interdisciplinary partnerships, and innovative technologies that drive bioinformatics research forward, paving the way for novel discoveries and transformative applications in agriculture research.

Dr. Lingling Jin, University of Saskatchewan

Genome Evolution and Reconstruction of Ancestral Monoploid Karyotypes

Evolution is a process that involves an organism's DNA changing over time. Often, this process is driven by external stressors (such as environmental changes, diseases, pests) and/or selective breeding. Recognizing how genomes evolve is of universal biological interest about species origin, survival and adaptation. Advances in computational infrastructure and computing power, coupled with the fast-increasing number and diversity of sequenced genomes, open the door to analyzing the dynamics of evolution at whole-genome resolution.

By using a successive ancestral contig construction technique, we have developed methods and a complete workflow to reconstruct chromosome-level ancestral genomes. The methods were applied and verified in many lineages across different studies that improve the scientific knowledge of genome evolution in various biological systems [1-8]. In the presentation, we will first introduce RACCROCHE [3], a pipeline for ancestral genome reconstruction, and our research on genome evolution based on genes, their syntenic orders, chromosomal gene co-occurrences, subgenome dominance in gene expressions [9], novel genomic variations [10] (structural variants and transposable elements) and their relationship with plant phenomics [11,12].

A particularly important application of this work is to crop species that exhibit rapid rates of evolution. By understanding crop evolution and its impact on plant phenotypes, breeders will be able to make informed decisions on how to breed crops, one of Canada's key economic resources.

Perl. In 2019, Dr. Wasmuth has appointed as Director of the Graduate College, a pan-Faculty community that strives to add value and leadership experience for UCalgary graduate students and postdocs.

Dr. Elisabeth Richardson, Mount Royal University

Comparative genomics of a pan-ciliate dataset reveal paradoxical protein complexes associated with membrane trafficking and peroxisome formation

The ciliate phylum is a group of protists noted for their unusual membrane trafficking system and apparent environmental ubiquity; as highly successful microbial predators, they are found in all manner of environments and the ability for specific species to adapt to extremely challenging conditions makes them valued as environmental bioindicators. We carried out a comparative genomic survey of selected membrane trafficking proteins in a pan-ciliate transcriptome and genome dataset. We observed considerable loss of membrane trafficking system (MTS) proteins that would indicate a loss of machinery that is generally conserved across eukaryotic diversity, even after controlling for potentially incomplete genome representation. In particular, the complete DSL1 complex was missing in all surveyed ciliates. This protein complex has been shown as involved in peroxisome biogenesis in some model systems, and a paucity of DSL1 components has been indicative of degenerate peroxisome. However, Tetrahymena thermophila (formerly Tetrahymena pyroformis) was one of the original models for visualizing peroxisomes. Conversely, the AP3 complex essential for mucocyst maturation in T. thermophila, is poorly conserved despite the presence of secretory lysosome-related organelles across ciliate diversity.

The 3rd Annual BioNet Conference Presentations

Title

A transcriptomics approach to understand the interactions between Escherichia coli *and copper, silver and gallium metal ions*

AUTHORS

Daniel A. Salazar-Alemán[1], Raymond J. Turner[1]

[1] University of Calgary, Calgary AB, Canada.

Presenting author: danielarmando.salaza@ucalgary.ca

KEYWORDS

Transcriptomics, copper, gallium, silver, *Escherichia coli*

ABSTRACT

In the current antimicrobial resistance era, where resistance to most antibiotics now common, metal and metalloid based antimicrobials (MBA) have re-surfaced as an alternative to handle infectious agents. A variety of MBAs are under research for their antimicrobial activities, with silver and copper products already in the market. However, the precise mechanisms of action behind them are not fully established. Here we explore the cell response profile of *Escherichia coli* when challenged to grow under the presence of silver nitrate, copper sulfate and gallium nitrate. Growth curves using a gradient of metal salt concentrations were run to determine a subinhibitory concentration for each of Ag+1 (10 µM), Cu+2 (39 µM) and Ga+3 (1250 µM). The chronic challenge took place in M9 minimal media + glucose and 1 x 10^9 cells were pelleted after 10 hours to perform total RNA extraction. After rRNA depletion and cDNA libraries creation, RNAseq was performed by means of Illumina MiSeq 150 (2.5M read pairs per sample, 3.8Gbp) with an overall >Q30 of 94.85%. Bioinformatic processing of data is underway using the rsubread and DESeq2 Bioconductor packages in R-Studio, and preliminary findings from the data analysis are expected to be presented. This will complement previous findings from a series of toxicogenomic screenings performed by our group years before [1-3].

REFERENCES

1. https://doi.org/10.1093/mtomcs/mfab071
2. https://doi.org/10.3390/genes10010034
3. https://dx.doi.org/10.3390%2Fgenes9070344

Genetic Analysis of Pediatric Obsessive-Compulsive Behaviors: the Effect of Genetic Variants on Imaging Endophenotypes

AUTHORS
Lilit Antonyan[1], S-M Shaheen[1], David R Rosenberg[2], Gregory L Hanna[3], Paul D Arnold[1]

[1]University of Calgary, Calgary AB, Canada
[2]Wayne State University, Detroit, MI, USA
[3]University of Michigan, Ann Arbor, MI, USA

Presenting author: lilit.antonyan1@ucalgary.ca

KEYWORDS
OCD, Genetics, Quantitative, Endophenotype, Neuroimaging

ABSTRACT
Background: Obsessive-compulsive (OC) behaviors (OCB) are characterized by intrusive thoughts and repetitive behaviors, which lie along a continuous distribution with clinical OCD at the upper extreme, consistent with the Research Domain Criteria (RDoC) paradigm. Also consistent with the RDoC model, the pathophysiology of OC traits involves structural and functional alterations within cortico-striato-thalamo-cortical (CSTC) circuits measured using neuroimaging. The goal of this study was to determine the relationship between brain structure and activity and childhood OCB within the RDoC paradigm.
Methods: All participants were assessed for OCD using a quantitative symptom measure: the Obsessive-Compulsive Scale of the Child Behavior Checklist Scale (CBCL-OCS). Around 8 million single nucleotide polymorphisms (SNPs) were genotyped and imputed on 863 non-related subjects. Genome-wide association study (GWAS) was performed. Later, SNP-based, and gene-based analyses were conducted. Genome-wide significant variants that are differentially expressed in brain tissue are identified. Next, polygenic risk score (PRS) analysis will be performed. Further, imaging data including cortical thickness and myelin water fraction (MWF) estimates were collected and analyzed on a subset of ~200 individuals who participated in our genetic study.
Results: 627 samples with CBCL-OCS scores, and 8 million SNPs passed the quality control filters. Although GWAS was underpowered, six variants from non-coding region of the 4q13.1 locus passed the p-value threshold ($5x10^{-8}$). The closest annotated gene is ADGRL3. After gene-based and gene-set enrichment analysis no specific genes were identified. After analysing available imaging measures, corpus callosum genu showed a significant correlation (P=0.029) between CBCL-OCS and MWF measures.
PRS and further imaging analysis are in progress.
Conclusions: Quantitative GWAS showed that 4q13.1 locus and corresponding haplotype could potentially be associated with the obsessive-compulsive symptom severity measured by CBCL-OCS. These findings, if they remain significant in a larger sample would provide insight to the pathophysiology of OC traits mediated by structural and functional changes in specific brain regions. To our knowledge, this project will be the largest to date to report genetic markers, and imaging endophenotypes of susceptibility to OCB in a pediatric-clinic-based population.

REFERENCES

1. https://pubmed.ncbi.nlm.nih.gov/8083134/
2. https://pubmed.ncbi.nlm.nih.gov/15335349/
3. https://dsm.psychiatryonline.org/doi/book/10.1176/appi.books.9780890425596
4. https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc
5. https://pubmed.ncbi.nlm.nih.gov/24840803/
6. https://pubmed.ncbi.nlm.nih.gov/17701901/
7. https://pubmed.ncbi.nlm.nih.gov/22820512/
8. https://pubmed.ncbi.nlm.nih.gov/29132412/

Power to detect episodic fitness shift at a small number of sites can be rescued with a covariate-informed branch-site type method

AUTHORS
Mobina Kazemi Mehrabadi[1,*], Kabita Baral[1,*] and A.P. Jason de Koning, Ph.D.[1]

[1]Department of Biochemistry and Molecular Biology, University of Calgary, Calgary AB, Canada and Alberta Children's Hospital Research Institute (ACHRI), Calgary AB, Canada

*authors contributed equally

Presenting author: Kabita.baral1@ucalgary.ca

KEYWORDS

Evolution, Positive Selection, fitness shift, likelihood ratio, codon.

ABSTRACT

Episodic positive selection (EPS) is generally believed to occur at a few sites and in a limited number of branches at a time, making it difficult to detect statistically. Branch-site tests are widely-used likelihood ratio tests that detect EPS along a prespecified lineage of a phylogeny at an unknown subset of sites. A limitation of power studies of branch-site methods has been a focus on detecting EPS simultaneously affecting an unrealistically large number of sites (*e.g.,* >20%). We simulated sequence alignments under a set of fitness shift scenarios using a non-equilibrium time-heterogeneous mutation-selection codon model at a small number of sites (~1%). Then, we evaluated the power of different branch-site tests to detect signatures of episodic adaptation. Our results demonstrated that these tests have very low power to identify signatures of fitness shift at a small number of sites. We then implemented a covariate-based method to improve the identifiability of positive selection by adding site-specific data informing the probability of mixture class assignments to sites. This enhanced statistical power by a factor of 10 when covariate information was reliable (for the best model, power was 32.48%). When the results were filtered by the Jensen Shannon divergence (JSD) between ancestral and derived stationary distributions, we found that the covariate method can improve power drastically (up to 96% power for modestly large fitness shifts). Without covariate information, none of the traditional branch-site tests appear to have appreciable power, even in the presence of large fitness shifts simultaneously affecting multiple sites.

Ancestral eukaryotic protein domain profiles for evolutionary cell biology

AUTHORS
Lael D. Barlow[1], Nathan A. Barlow[2], and Mark C. Field[1,3]

[1] Division of Biological Chemistry and Drug Discovery, School of Life Sciences, University of Dundee, Dundee, UK.
[2] No affiliation.
[3] Institute of Parasitology, Biology Centre, Czech Academy of Sciences, 37005 České Budějovice, Czech Republic.

Presenting author: lael@ualberta.ca

KEYWORDS
Microbial eukaryotes, molecular evolution, protein family, phylogeny, orthology

ABSTRACT
Eukaryotes diversified over 1 billion years ago, making the systematic comparison of cell biology among extant eukaryotes, including livestock, crops, and pathogens, a complex task that requires tracing molecular evolution over long periods of divergence. Despite increasing genomic and transcriptomic data availability, classifying numerous eukaryotic proteins into reliable orthologous groups remains challenging. Hence, whether homologous proteins possess similar or contrasting functions is often unclear. To address this issue, we have generated a highly selective database of protein domain profiles (sequence alignments) by excluding phylogenetically redundant or uninformative sequences and aligning the remaining sequences with probable orthology. This database enables easy querying with eukaryotic protein sequences using sensitive profile-profile alignment methods, rapidly generating phylogenetically informative alignments. The application of this database will facilitate the understanding of molecular evolution and functional relationships among eukaryotic proteins in model and non-model species, ultimately aiding comparative cell biology and the development of targeted strategies for improving livestock, crop production, and pathogen control.

A statistical method for image-mediated association studies discovers genes and pathways associated with four brain disorders

AUTHORS

Jingni He[1], Lilit Antonyan[2,3], Harold Zhu[4], Qing Li[1], David Enoma[1], William Zhang[5], Andy Liu[6], Bo Cao[7], M. Ethan MacDonald[3,8,9,10,12], Paul D. Arnold[2,3,11,12,] and Quan Long[1,2,3,12,13]

[1] Dept of Biochemistry and Molecular Biology, Cumming School of Medicine, University of Calgary
[2] Dept of Medical Genetics, Cumming School of Medicine, University of Calgary
[3] Hotchkiss Brain Institute, Cumming School of Medicine, University of Calgary
[4] Dept of Biological Sciences, Faculty of Science, University of Calgary
[5] The Harker School, San Jose, CA, USA
[6] Sir Winston Churchill High School, Calgary
[7] Dept of Psychiatry, Faculty of Medicine & Dentistry, University of Alberta
[8] Dept of Biomedical Engineering, Schulich School of Engineering, University of Calgary
[9] Dept of Electrical and Software Engineering, Schulich School of Engineering, University of Calgary
[10] Dept of Radiology, Cumming School of Medicine, University of Calgary
[11] Dept of Psychiatry, Cumming School of Medicine, University of Calgary
[12] Alberta Children's Hospital Research Institute, Cumming School of Medicine, University of Calgary
[13] Dept of Mathematics and Statistics, Faculty of Science, University of Calgary

Presenting author: David Enoma (david.enoma@ucalgary.ca)

KEYWORDS

Image-Mediated Association Study, Image-Derived Phenotypes, GWAS, Cerebellar Abnormalities, Neuropsychiatric Disorders

ABSTRACT

[Background] Brain imaging and genomic sequencing are powerful tools in studying the genetic basis underlying brain disorders such as schizophrenia (SCZ), major depression disorder (MDD), bipolar disorder (BPD), and autism spectrum disorder (ASD) [1]. However, imaging large cohorts is expensive and may not be available in legacy genome-wide association study (GWAS) datasets. Fortunately, the UK Biobank database has collected genotype and phenotype data, including nearly 50,000 volunteers with brain MRI scans [2], [3]. T1 and T2 weighted images, & diffusion RI are among the contrast types used; neuroimaging endophenotypes called image-derived phenotypes (IDPs) are also provided. Using an integrated feature selection/aggregation model, we developed Image-Mediated Association Study (IMAS) using imaging/genomics data to conduct association mapping in legacy GWAS cohorts. We leverage our improvement on the Transcriptome-Wide Association Study (TWAS) framework [4] by data bridging through disentangling TWAS into feature selection and aggregation [5].

[Methods] In the integrated feature selection/aggregation model, IMAS utilizes borrowed imaging/genomics data to conduct association mapping in legacy GWAS cohorts. We leveraged the UK Biobank IDPs with IMAS to discover the genetic basis underlying SCZ, MDD, BPD & ASD and verified them by analyzing annotations, pathways and expression quantitative trait loci (eQTLs). For feature selection, we implemented two models: regularized regression (Elastic Net[4]) and linear mixed model

(EMMAX[6],[7]). Consequently, we implemented multiple linear regression (original TWAS) [4] and a weighted kernel as implemented in SKAT [8]. We performed a simulation based on the real UK Biobank genotype, and using the 1,000 Genomes data, we did type-I error estimation and power calculations. We collected 33,553 individuals in UK Biobank for 885 IDPs and performed IMAS 2,548, 3,644, 1,028, and 7,068 individuals with SCZ, MDD, BPD, and ASD, respectively. We conducted pathway enrichment analysis, Standard GWAS, and eQTL analysis and developed a Gene-to-gene connection network and cross-disorder co-expression matrix of cerebellum-related genes.

[Results & Impact] IMAS successfully discovered IDPs associated with the neuropsychiatric disorders under study and SNVs enriched in biologically relevant pathways. Our method delineates individual and shared genetic mechanisms underlying cerebellar abnormalities shared by four disorders. The top SNVs overlapped with eQTLs important for neuropsychiatric disorders. Standard GWAS cannot discover the IMAS-selected SNVs and pathways. IMAS is theoretically more powerful than a hypothetical protocol with imaging available in the GWAS dataset. We show the feasibility of reanalyzing legacy GWAS datasets without conducting additional imaging, bringing cost savings for integrated analysis of genetics, imaging, and brain disorders.

REFERENCES

[1] P. Falkai, A. Schmitt, and N. Andreasen, “Forty years of structural brain imaging in mental disorders: is it clinically useful or not?,” Dialogues in Clinical Neuroscience, vol. 20, no. 3, pp. 179–186, Sep. 2018, doi: 10.31887/DCNS.2018.20.3/pfalkai.
[2] L. T. Elliott et al., “Genome-wide association studies of brain imaging phenotypes in UK Biobank,” Nature, vol. 562, no. 7726, Art. no. 7726, Oct. 2018, doi: 10.1038/s41586-018-0571-7.
[3] N. Rusk, “The UK Biobank,” Nat Methods, vol. 15, no. 12, Art. no. 12, Dec. 2018, doi: 10.1038/s41592-018-0245-2.
[4] E. R. Gamazon et al., “A gene-based association method for mapping traits using reference transcriptome data,” Nat Genet, vol. 47, no. 9, Art. no. 9, Sep. 2015, doi: 10.1038/ng.3367.
[5] C. Cao et al., “Disentangling genetic feature selection and aggregation in transcriptome-wide association studies,” Genetics, vol. 220, no. 2, p. iyab216, Feb. 2022, doi: 10.1093/genetics/iyab216.
[6] H. M. Kang et al., “Variance component model to account for sample structure in genome-wide association studies,” Nat Genet, vol. 42, no. 4, Art. no. 4, Apr. 2010, doi: 10.1038/ng.548.
[7] A. C. Samson, R. F. Dougherty, I. A. Lee, J. M. Phillips, J. J. Gross, and A. Y. Hardan, “White matter structure in the uncinate fasciculus: Implications for socio-affective deficits in Autism Spectrum Disorder,” Psychiatry Research: Neuroimaging, vol. 255, pp. 66–74, Sep. 2016, doi: 10.1016/j.pscychresns.2016.08.004.
[8] M. C. Wu, S. Lee, T. Cai, Y. Li, M. Boehnke, and X. Lin, “Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test,” The American Journal of Human Genetics, vol. 89, no. 1, pp. 82–93, Jul. 2011, doi: 10.1016/j.ajhg.2011.05.029.

The Evolution of pathogen testing and reporting during a pandemic: Incorporating bioinformatics and high throughput sequencing into public health surveillance

AUTHORS

, Kara Gill[a], Kanti Pabbaraju[a], Anita A. Wong[a], Paul Dieu[a], Matthew Croxen[b,c,e], Graham A. Tipples[b,e,f], Tarah Lynch[a,d]

[a] Alberta Precision Laboratories, Public Health Laboratory, Calgary, Alberta, Canada
[b] Alberta Precision Laboratories, Public Health Laboratory, Edmonton, Alberta, Canada
[c] Department of Laboratory Medicine and Pathology, University of Alberta, Edmonton, Alberta, Canada
[d] Department of Pathology and Laboratory Medicine, University of Calgary, Calgary, Alberta, Canada
[e] Li Ka Shing Institute of Virology, University of Alberta, Edmonton, Alberta, Canada
[f] Department of Medical Microbiology and Immunology, University of Alberta, Edmonton, Alberta, Canada

Presenting author: kara.gill@aplabs.ca

KEYWORDS

Real-time RT-PCR, genome sequencing, SARS-CoV-2, quality control, public health, surveillance

ABSTRACT

When COVID-19 caused by SARS-CoV-2 emerged as a global pandemic, the Alberta Precision Laboratories, Public Health Laboratory (ProvLab) responded by rapidly developing and implementing nucleic acid based detection assays. Real-time reverse transcriptase PCR (rtRT-PCR) assays are sensitive, specific, reproducible, high-throughput and provide a fast turnaround time. The first iterations of these assays were used for the detection of SARS-CoV-2, however, as variants of concern (VOC) emerged, ProvLab developed multiple rtRT-PCR assays targeting the discriminatory mutations for each VOC allowing for the accurate identification and tracking of these emerging threats.

During this time, ProvLab was also validating genome sequencing (GS) methods using Oxford Nanopore and Illumina sequencing technologies. This led to the development of a combined approach for reporting, where rtRT-PCRs were used to rapidly detect VOCs, and genome sequencing was used for a thorough characterization of the genome to continuously monitor for evolving mutations and novel variants. As the pandemic has progressed, ProvLab has focused on decreasing the turnaround time for genome sequencing by streamlining the wet-lab and bioinformatic workflows and moved to report SARS-CoV-2 lineages based solely on GS since March 2023.

To achieve this shift in reporting, ProvLab was required to develop additional quality control measures for GS data. Lab-developed PCR assays require rigorous validation prior to testing and reporting patient samples including evaluation of sensitivity, specificity, accuracy, reproducibility, efficiency, linearity, and limit of detection. Elements such as positive and negative controls for monitoring assay performance and detecting contamination are shared between PCR and GS. However, the quality and quantity of data produced with GS requires an alternate set of parameters to ensure accurate and reliable results. These include metrics such as read quality, percentage of mapped reads, mapping quality scores, coverage

distribution, and sequencing depth. These parameters are critical for producing high-quality GS data that can be used to effectively track and monitor viral evolution.

Bioinformatics-driven genomic surveillance can help monitor the spread of a virus and identify new variants, thereby aiding in the development of targeted public health measures. This combined approach of rtRT-PCR and GS has proven to be a useful laboratory algorithm, allowing for rapid detection of VOCs and providing essential, high-quality information on the virus's ongoing evolution and transmission dynamics.

REFERENCES

1. https://doi.org/10.3138/jammi-2020-0026
2. https://doi.org/10.1128/Spectrum.00315-21
3. https://doi.org/10.1128/JCM.01110-20

Unraveling the mechanism of *ToxB* replication in *Pyrenophora tritici-repentis*

Ryan Gourlie[1,2], Megan McDonald[3], Dmytro Yevtushenko[2], and Reem Aboukhaddour[1]

[1]Agriculture and Agri-food Canada, Lethbridge, Alberta, Canada
[2]University of Lethbridge, Lethbridge, Alberta, Canada
[3]University of Birmingham, Birmingham, United Kingdom

Presenting author: ryan.gourlie@agr.gc.ca

Keywords
plant pathogens, fungal effectors, transposons, virulence evolution, comparative genomics

Pyrenophora tritici-repentis (Ptr) is a fungal pathogen that causes tan spot of wheat. This pathogen secretes several necrotrophic effectors, including the protein ToxB, which induces chlorosis. The *ToxB* gene exists in multiple copies, ranging from 1 to 10 in different isolates. Some isolates may carry an inactive homolog called *toxb*, which is present as a single copy. Previously, we have shown that *ToxB* is present on the putative Starship transposon *Icarus* [1]. However, the presence of *ToxB* within *Icarus* did not explain its multicopy nature, and further investigation was needed. To understand the mechanism by which the *ToxB* gene was replicating, a set of eight long-read sequenced (PacBio RS II) Ptr genomes (CANU assembled) were used for comparative analysis to define the replication unit. The results showed some structural features that support the presence of a transposon. The size of the putative transposon varies from 5,800 bp up to 18,500 bp, likely due to other transposon insertions. Different methods of alignment produced consistent edges which appear to be unmatching sequences (i.e. not LTR or TIR). Within these edges was a predicted ORF containing domains associated with reverse transcriptase and RNase H type-1 proteins (UniProt search), both of which are associated with Class I retrotransposons. A BLAST search of RepBase (transposons database) for this gene, showed some similarity to *I-2_CH* (67% identity over 96% of the query) which is a non-LTR retrotransposon identified in the necrotrophic maize pathogen *Cochliobolus heterostrophus* [2]. Additionally, in many instances, very near the putative edge lies an integrase gene that is required for the insertion of retrotransposons, suggesting that the transposon may be non-autonomous. Interestingly, in a non-pathogenic isolate, inactive *toxb* appears to be located just downstream of the putative transposon, lying outside the defined edges. The search for a target site duplication is ongoing, and should it be found, it would confirm these findings as a transposon. To aid with this search, sequencing of more isolates, including closely related species in the *Pyrenophora* genus is underway. The evolutionary history of multi-copy *ToxB/toxb* is complicated and difficult to unravel. This research marks a significant step forward in our understanding of this virulence gene.

REFERENCES

[1] https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-022-01433-w
[2] https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-536

Shining light on circRNA dysregulation in neurodegenerative disease

AUTHORS
Travis Haight [1,2], Madison Bronk [1,2], Riya Roy [1,2], Athanasios Zovoilis [1,2]
[1] University of Lethbridge, Lethbridge AB, Canada
[2] Southern Alberta Genome Science Center (SAGSC)

Presenting author: haight@uleth.ca

KEYWORDS
Circular RNA, non-coding RNA, neurodegenerative disease, RNA dysregulation, Next-Generational Sequencing

ABSTRACT
Circular RNAs (circRNA) are a more recent addition to the class of non-coding RNA (ncRNA) in which they are well known for their stability due to their lack of a 5' m7g cap or 3' poly-A tail leading to a natural protection from various endogenous exonucleases. The rapidly and ever-expanding world of research into circRNA biogenesis and function implies that they are simply not splicing errors as previously thought but hold conserved biological functions contributing to regulation of gene expression. Here we have employed and developed an integrative RNA genomics approach to study circRNA which utilizes Ilumina short read sequencing data, and Nanopore long read sequencing to determine how circRNA expression changes in response to a neurodegenerative disease state. We have focused on mouse tissue and cell line models of amyloid pathology.

AcrTransAct: Anti-CRISPR Transformer-based Activity Prediction

AUTHORS

Moein Hasani[1], Chantel Trost[2], Lingling Jin[1]
[1] Department of Computer Science, University of Saskatchewan, Saskatoon, SK, S7N 5C9, Canada
[2] Department of Molecular Genetics, University of Toronto, Toronto, ON, M5S 1A8, Canada

Presenting author: moein.hasani@usask.ca

KEYWORDS

Anti-CRISPR, CRISPR-Cas, Transformers, Deep Learning, Neural Networks

ABSTRACT

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) [1] and CRISPR-associated (Cas) proteins provide a powerful and efficient mechanism for bacterial immunity against foreign DNA such as bacteriophages. The Red Queen hypothesis proposes that organisms must continually evolve new mechanisms of resistance to parasites to avoid extinction; therefore, some bacteriophages and mobile genetic elements that infect bacteria have evolved anti-CRISPR (Acr) [2] proteins that inhibit CRISPR-Cas systems, enabling their own survival. Since CRISPR-Cas systems can precisely target DNA, they have been harnessed for biotechnology and medical applications. However, because Acrs disable or modulate CRISPR-Cas activity, it is vital that we understand how they work. Unfortunately, the experimental detection of Acr activity can be time-consuming and costly, and protein-protein interaction software cannot predict Acr-mediated CRISPR-Cas inhibition due to the complex interactions of Acrs with multiple Cas proteins, especially for Class I CRISPR-Cas complexes. To solve this problem, we present AcrTransAct, which predicts the inhibition of CRISPR-Cas systems by Acrs using transformer-based Deep Neural Networks (DNN). We created an inhibition dataset compiled from two online Acr databases, AcrHub and Anti-CRISPRdb. We designed a DNN that was trained on this dataset, which can predict the likelihood of an Acr inhibiting a CRISPR-Cas system. Our classification network is a Convolutional Neural Network (CNN) that uses 1-D convolution layers and receives features extracted by the Evolutionary Scale Modeling (ESM) transformer protein model [3] and structural information from the NetSurfp-3.0 [4] as input. We achieved a 5-fold cross-validation accuracy of 84% and F1 score of 0.85 in predicting the inhibition of I-C, I-E, and I-F CRISPR-Cas systems by Acrs in our dataset. By comparing the results of feeding the ESM features to the classification network against those obtained by feeding one-hot encoded sequences and structural information, we demonstrate the richness of information extracted by the ESM network from protein sequences. Our web application enables users to submit their Acr protein sequences and receive a comprehensive report on the likelihood of inhibition of multiple CRISPR-Cas systems by the provided Acr sequence. Our work provides a valuable tool for predicting interactions between Acrs and CRISPR-Cas systems and facilitates experimental Acr activity experiments by selecting the most likely Acr from many homologous candidate proteins. Furthermore, we provide insights into the capabilities of transformer networks in biological sequence analysis tasks, especially in the context of protein-protein interactions.

Inference of subgenomes resulting from polyploid events using synteny based dynamic linking and maximum neighbourhood

AUTHORS

Thulani Hewavithana[1], Chu Shin Koh[2], Avleen Kaur[1], Raymond J. Spiteri[1], Isobel Parkin[3*], and Lingling Jin [1*]

[1] Department of Computer Science, University of Saskatchewan, Saskatoon, SK, S7N 5C9, Canada
[2] Global Institute for Food Security, SK, Canada
[3] Agriculture and Agri-Food Canada, Saskatoon RDC, SK, Canada
Presenting author: qnm481@usask.ca

KEYWORDS
Synteny, Subgenomes, Gene fractionation, Polyploidy, Collinear blocks, Homologous genes

ABSTRACT
Polyploidy is a common occurrence in flowering plants, where whole genome duplication and triplication events result in additional sets of chromosomes (or subgenomes) [1]. Almost all flowering plants have undergone at least one polyploid event, with some experiencing multiple events since their ancestral angiosperm. Synteny blocks, i.e., contiguous regions of genomic DNA sharing gene content and order, aid in inferring the evolutionary footprint of the genomes. Polyploidy leads to subgenomes with redundant gene copies that are rapidly lost through gene fractionation. Assigning synteny blocks to different subgenomes (representing subsets of the organism's genome that have undergone polyploid events) is challenging due to recurring polyploidization and fractionation events. These events complicate the situation by scrambling gene order on a background of evolutionary processes such as gene family expansion, gene loss, and genome rearrangement. Currently, there is no automated subgenome reconstruction method, and the existing methods [2] require manual curation. To address this challenge, we developed the "SyntenyLink" algorithm that automatically reconstructs subgenomes from synteny blocks. The algorithm considers differences showed in substitution and fractionation patterns in synteny blocks, as well as continuity of conserved order of genes to reconstruct the most parsimonious subgenomes.
The algorithm first utilizes the blastp [3] program and DAGchainer [4] to identify synteny blocks across different chromosomes of two related genomes. It then organizes the blocks into subgenomes using depth-first search on a weighted graph where the vertices in the graph represent super synteny blocks identified by translocation breakpoints. The graph edges are weighted using the combined information of percent identity, block chain, and gene density between the two vertices connected by the edge. The algorithm then minimizes the number of translocation events by using a maximum neighborhood method.

The SyntenyLink algorithm has been validated and tested using published subgenomes of Brassica rapa [2]. Two further case studies conducted on different plant species, Triticum turgidum sp. and Vitis vinifera, illustrate the algorithm's potential applications in polyploid genomes in other orders of the angiosperm tree. The results from Brassica rapa demonstrate that the SyntenyLink algorithm can achieve an accuracy of 88% in reconstructing subgenome 1 and 78% in reconstructing subgenomes 2 and 3.

This algorithm represents a promising tool for reconstructing subgenomes from complex polyploid genomes, with far-reaching implications for the study of the evolutionary history of flowering plants and other polyploid organisms.

Comparative Evaluation of Transfer Learning Models for Accurate Classification of Septoria Tritici Blotch and Strip Rust in Wheat Leaves

Yuanyuan Ji, Lingling Jin
Department of Computer Science, University of Saskatchewan, Saskatoon, SK, S7N 5C9, Canada
Presenting author:yuj011@usask.ca

KEYWORDS

Wheat, Strip rust, Septoria, phenotyping, Deep learning, Transfer learning

ABSTRACT

Deep learning has achieved great success in classifying over 1 million images with significantly increased accuracy on various public datasets (ImageNet, CIFAR-10) compared to traditional machine learning algorithms. The field of plant pathology holds great promise for revolutionizing high-throughput field phenotyping using deep learning algorithms. Wheat is one of the most important crops in the world, accounting for 20% of the calories consumed by humans [1]. However, over 80% and 50% of the area of wheat cultivation is susceptible to Stripe rust (*Puccinia striiformis*) and Septoria tritici blotch (*Zymoseptoria tritici*), respectively [2]. Accurate identification of fungal pathogens from the symptoms exhibited on wheat leaves is crucial for breeders and farmers because it can help them understand the distribution of the disease and then make better decisions on how to respond to the spread of the disease. This study performs a comprehensive evaluation of different transfer learning models in classifying fungal pathogens using images of wheat leaves. The performance of four deep learning models, ResNet50, Inception v3, VGG16, and Xception, was compared with a basic 3-layer Convolutional Neural Network (CNN). All models were trained on 407 images of wheat leaves in three categories, including Strip rust and Septoria blotch infected leaves and healthy leaves. The dataset was split into three subsets (70%, 15% and 15% ratio) for training, validation, and testing. Cross-entropy loss is used in all the transfer learning models to update the weights during the training process. Data augmentation was applied to balance the number of training data in each category. The study shows that the Inception v3 model achieved the highest accuracy of 99.0% on the test set compared to the CNN model with 95.83% accuracy, while other models showed relatively lower accuracy between 20% to 94.4%. The ResNet 50 and VGG16 models showed low accuracy, which could be related to field backgrounds making feature extraction difficult. To improve the accuracy of these two models, the images were preprocessed to reduce the background noise before feeding to the networks by cropping only the local scope of the infection area. After preprocessing, all the transfer learning models showed a promising performance with accuracies of over 90%.

REFERENCE

1. https://doi.org/10.1038/nature11650
2. https://doi.org/10.3389/fpls.2022.1034600

Representation learning and transfer learning: applications to genetics and omics

AUTHOR
Ariel Ghislain Kemogne Kamdoum [1]
[1] Dept of Mathematics and Statistics, Faculty of Science, University of Calgary, Calgary, Canada
Presenting author: arielghislain.kemogn@ucalgary.ca

KEYWORDS:
Representation learning, transfer learning, autoencoders, transcriptome-wide association study (TWAS)

ABSTRACT
[Background] The common challenge we face in medical genetics while analyzing -omics data is the problem of complexity and high dimensionality. Deep learning, via representation and transfer learning, help address these challenges by omitting unnecessary details and capturing relevant information from the data, improving predictions, and understanding of disease mechanisms. A simple example of representation learning is like observing a cat image and identifying it as a "Cat" by focusing on the face features while ignoring unnecessary details from the rest of its body. Basically, the pixels in an image of a face are the “default” representation however the actual objects (such as eyes and noses) are the “correct” representations. Representation learning simplifies data, retaining important content for various tasks[1] including transfer learning, which enhances performance across domains[6]. It can be applied to fields such as image processing[2], NLP[3], and genomics[4], assisting in biomarker identification[5] and discovering new genetic associations[9]. As an illustration, Google’s Enformer[11], a deep neural network using the Transformer architecture[13], excels at predicting enhancer-promoter interactions from DNA sequences. It learns DNA vector representations, captures complex genomic patterns, and integrates 5,313 human multiomicstracks and reference genomes. Although not tailored for specific diseases, our lab's TF-Enformer[14] refines it for breast cancer, improving performance by 7% and finding 19 new susceptibility genes.
[Aims & Methods] Representation learning, which is composed of multiple non-linear transformations aiming to create sensible representations, is very beneficial for high-dimensional data with complex nonlinear structures in fields like genetics and omics[8]. Hence, techniques such as autoencoders which generate sensible representations by encoding inputs, are extremely useful. Our research’s plan is to make use of such methods while exploring the capabilities of (non)linear models in predicting phenotypes from genomic data and investigate representation learning's role in enhancing their accuracy and stability. We plan to undertake the following projects: (1) Identifying when linear models can approximate nonlinear ones. (2) Developing a simple model to approximate nonlinear models, integrating prior biological information, and understanding representation learning's role.(3) Determining mathematical conditions for greater nonlinear model stability, while assessing conditions when a network is more stable than single genes in predicting phenotype. We aim to provide an in-depth analysis of representation and transfer learning, delving into their mathematical foundations, and demonstrate their application in TWAS and GWAS while examining the conditions for effective transfer learning implementation.
[Impact] Exploring nonlinear structures in representation learning enhances genetics and omics comprehension, supports accurate complex biological system modeling, and improves diagnoses, treatments, and genomics/proteomics data analysis tools.

REFERENCES

[1] Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives.

IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1798-1828.
[2] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.
[3] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
[4] Zitnik, M., Nguyen, F., Wang, B., Leskovec, J., Goldenberg, A., & Hoffman, M. M. (2019). Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. Information Fusion, 50, 71-91.
[5] Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., ... & Greene, C. S.
(2018). Opportunities and obstacles for deep learning in biology and medicine. Journal of The Royal Society
Interface, 15(141), 20170387.
[6] Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345-1359.
[7] Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? In Advances in Neural Information Processing Systems (pp. 3320-3328).
[8] Gamazon, E. R., Wheeler, H. E., Shah, K. P., Mozaffari, S. V., Aquino-Michaels, K., Carroll, R. J., ... & Im, H.
K. (2015). A gene-based association method for mapping traits using reference transcriptome data. Nature
Genetics, 47(9), 1091-1098.
[9] Wainberg, M., Sinnott-Armstrong, N., Mancuso, N., Barbeira, A. N., Knowles, D. A., Golan, D., ... & The
GTEx Consortium. (2019). Opportunities and challenges for transcriptome-wide association studies. Nature
Genetics, 51(4), 592-599.
[10] Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. Journal of Big Data, 3(1), 9.
[11] Z. Avsec et al., Effective gene expression prediction from sequence by integrating long-range interactions, (in English), Nat Methods, vol. 18, no. 10, pp. 1196-+, Oct 2021, doi:10.1038/s41592-021-01252-x.
[12] W. Q. Wen et al., Genetic variations of DNA bindings of FOXA1 and co-factors in breast cancer susceptibility, (in English), Nat Commun, vol. 12, no. 1, Sep 13 2021, doi: ARTN 531810.1038/s41467-021-
25670-9.
[13] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).
[14] Li Q, Chen Z, Wen W, Wang D, Yan J, Shu X, Zheng W, Guo X#, Long Q#. Leveraging deep transfer learning and multi-omics integration to enhance regulatory variant prediction. To be submitted to Nature
Methods in June 2023. * = co-first; #= co-corresponding;

Fungal communities are suitable soil health indicators in agricultural lands in Alberta.

Nafsa Khazaei [1], Brian Lanoil [1], Derek Mackenzie [2]

1. Department of Biological Sciences, University of Alberta, Edmonton, AB, T6G2R3
2. Department of Renewable Resources, University of Alberta, Edmonton, AB, T6G2R3
Presenting author: nafsa@ualberta.ca

KEYWORDS

Soil health, agricultural practices, ITS gene, fungal diversity, functional guilds

ABSTRACT

In 2019-2020, Alberta soil quality monitoring program (SQMP) sampled thirty-eight benchmark farm sites to examine the effects of different agricultural practices on soil health incorporating soil microbial communities. At each site, surface (0-10 cm) soils were collected from triplicate samples at each of three different field positions: upper, middle, and lower. We investigated the soil fungal community through high-throughput sequencing of the ITS gene and evaluated the response of the fungal community to tillage intensity, herbicide use, fertilization methods, and crop types. High and no tilled soils as well as crops such as canola, wheat and barley showed the highest alpha diversity. Low and no-till samples, wheat crop, fallow farms with no herbicide, and no fertilizer samples showed the highest levels of heterogeneity. In all levels of tillage, herbicide usage and fertilization systems, arbuscular mycorrhizal fungi (AMF) and wood and dung saprotrophic fungi were the most abundant functional guilds; however, there was noticeable variation in the abundance of plant pathogen and endophytic functional guilds in these groups of agricultural practices. Different crop types varied in the abundance of most functional guilds including AMF, wood and dung saprotroph, ectomycorrhizal fungi, endophyte, lichenized, and plant pathogenic fungi.

Overall, increased fungal diversity and the abundance of specific fungal functional guilds were sensitive to agricultural management practices. As a result, we conclude that genomic data of soil fungal communities could be a potential tool for soil health assessment.

Decoding genomic adaptations of an alkaliphilic Opitutales using gene-tree species-tree reconciliation

AUTHORS
Damon Mosier[1], Varada Khot[1], Dmitry Y Sorokin[2], Marc Strous[1]

[1] University of Calgary, Calgary, Canada
[2]TU Delft, Delft, Netherlands

Presenting author: vmkhot@ucalgary.ca

KEYWORDS

Ancestral genome reconstruction, evolution, comparative genomics, phylogenetics, soda lakes

ABSTRACT

On a planet with such a diverse range of environments, it is intriguing to discover a bacterium from the order level that is so adaptable. Members from the Opitutales order have been found in freshwater, marine, sediment, and insect gut habitats, with vastly varying genomes ranging in size from 2Mbp to 7.5Mbp. In this study, we dive into the evolutionary history of a novel Opitutales bacterium isolated from the alkaline soda lakes of the Kulunda Steppe in Russia, using ancestral genome reconstruction and phylogenetics. Our workflow [1] involved creating gene ortholog groups of related Opitutales members and conducting a gene-tree species-tree reconciliation [2] to investigate the hypothesized adaptation of this organism from a marine to an alkaline environment. Surprisingly, our results suggested that the soda lake Opitutales most likely evolved first. We continue to dive into what genes were transferred, lost, and duplicated during the various environmental adaptations of this cosmopolitan organism and when these events occurred. Moveover, what unique genomic features define the signature of the alkaline Opitutales.

REFERENCES

1. https://link.springer.com/protocol/10.1007/978-1-0716-2691-7_9
2. https://royalsocietypublishing.org/doi/full/10.1098/rstb.2014.0335

Gene-Environment Interactions: Peer Victimization Experience and Genetic Risk for Depression

AUTHORS
Min Jae Kim[1], S-M Shaheen[1], Heather Brittain[2], Tracy Vaillancourt[2], Paul Arnold[1]
[1] University of Calgary, Calgary, Canada
[2] University of Ottawa, Ottawa, Canada

Presenting author: minjae.kim@ucalgary.ca

KEYWORDS
Genetics, Mental Health, Depression, Peer Victimization, Environmental Influences on Health, Pediatrics

ABSTRACT
Adverse childhood experiences such as peer victimization (i.e., bullying) is a common phenomenon, where the experience can lead to an increased susceptibility to various psychiatric disorders including depression. However, the experience of peer victimization has varying effects between individuals, and therefore further research is required to understand how genetic predisposition in conjunction with environmental factors interacts to confer risk for depression. Gene-by-environment (G x E) interaction studies on depression have been conducted in the past with a focus on candidate genes, but findings have been inconsistent. Recent genetic studies have identified the effect of polygenic risk on the disorder, where polygenic risk scores (PRS) obtained from genome-wide association studies (GWAS) can be used to predict the risk of developing depression. Testing both candidate genes and PRS, and their interaction with environmental factors may be a promising approach to understanding the complex aetiology of depression.

Longitudinal data from McMaster Teen Study have been obtained, where students initially assessed in Grade 5 have been followed up until the age of 22 (n=349). Candidate gene analysis has been conducted using polymerase chain reaction and gel electrophoresis on monoamine oxidase A gene (MAOA) and dopamine transporter gene (DAT1/SLC6A3) to determine the variable number tandem repeats (VNTRs) of each participant. The variation in VNTRs will be tested to determine their association with depressive symptoms. Genome-wide association study will be conducted to identify the genetic variants associated with depression and to calculate the polygenic risk score (PRS) of each participant's susceptibility to developing depression. History of childhood peer victimization will be included along with candidate gene analysis and estimated PRS data to investigate the gene-by-environment interactions associated with depressive symptoms.

Through the findings, both candidate genetic variants and their polygenic risk scores may potentially explain individual differences in the development of psychiatric disorders following adverse environmental exposures.

REFERENCES
1. https://www.ncbi.nlm.nih.gov/pubmed/XXX
2. https://link.springer.com/article/
3. https://doi.org/

Taxonomic classification and analysis workflow for microbiome sequence data using Snakemake, R, and R markdown

AUTHORS
Lance Lansing[1], Marta Guarna[2], Rodrigo Ortega Polo[1]
[1] AAFC Lethbridge Research and Development Centre, Lethbridge, Canada
[2] AAFC Lacombe Research and Development Centre, Lacombe, Canada

Presenting author: Lance.Lansing@agr.gc.ca

KEYWORDS
Metagenomics, taxonomic classification, bioinformatics methods, microbiome, workflows

ABSTRACT
Metagenomic sequence data is an increasingly common subject of biological research and many research questions necessitate the identification of the microbial community. Taxonomic classification is required for such measurements and analyses as community composition, alpha and beta diversity, and differential abundance across of taxa across experimental treatments. In this presentation, we describe the workflow we've developed for producing and analyzing taxonomic classification data from microbiome shotgun sequences. We also highlight the technologies utilized for analysis, visualization, and reporting, including the Snakemake workflow management system, and the R language and related tools.

Our workflow begins with a Snakemake pipeline that takes as input raw microbial sequence reads, performing sequence trimming, quality control and reporting, and taxonomic profiling via Kraken2. This is followed by aggregation of Kraken2 classification reports, producing a matrix containing taxa counts for all samples. We use an SQLite database and R Shiny interface tool to store and access classification matrices from different experiments. Our R scripts apply a combination of analysis packages, including ANCOMBC and indicspecies for differential abundance analysis and vegan for beta and alpha diversity metrics. Finally, the tabular results and visualizations are automatically compiled into polished PDF reports using R Markdown.

We will describe the factors influencing our tool and method choices and assess the benefits and difficulties we've encountered. Our workflow has evolved over time with tools and technologies being added or replaced. As metagenomic methods continue to be created and enhanced, a flexible development process and workflow structure prove to be invaluable.

Development of Tools for the Interpretation of Influenza Next Generation Sequencing Analysis Data

AUTHORS
Naomi Ie[1], Xiaoli Dong[1], Linda Lee[1], Kanti Pabbaraju[1], Alex Presbitero[1], Tarah Lynch[12], Matthew Croxen[34]
[1] Alberta Precision Laboratories Public Health Laboratory South, Calgary, Canada
[2] University of Calgary, Calgary, Canada
[3] Alberta Precision Laboratories Public Health Laboratory North, Edmonton, Canada
[4] University of Alberta, Edmonton, Canada

Presenting author: naomi.ie@albertaprecisionlabs.ca

KEYWORDS
public health surveillance, R, influenza, next-generation sequencing, infectious disease, virology

ABSTRACT
Influenza viruses are single-stranded, segmented RNA viruses that can cause respiratory infections in a variety of species, including humans [1]. These viruses are highly infectious [1], can cause severe disease in vulnerable individuals [1], and are capable of mutating rapidly via genetic drift and reassortment [1-2]. As a result, these viruses are seen as a threat to public health [3], and it is important to develop methods to identify new circulating strains.
The Provincial Laboratory for Public Health has developed a pipeline for the analysis of influenza A and B Illumina sequencing data from clinical positive specimens. This pipeline uses a reference-based assembly approach, but due to the genetic diversity of influenza viruses, it is not possible to use one reference sequence for all strains of influenza A or B. Therefore, the analysis pipeline selects an appropriate reference sequence based on matches to raw reads. Furthermore, due to the potential for reassortment, a reference sequence is selected for each individual segment.
Many of the outputs from this analysis pipeline have separate data for each segment within each specimen, which can cause challenges in the interpretation of data from these outputs. Also, none of these outputs contain specimen information (eg. source, accession number, clinical results, preparation protocol), and difficulties have been encountered in attempts to add this information to the output using the tools available in Microsoft Excel, such as VLOOKUP formulas. R scripts were developed to accomplish both of the above tasks. One R script was developed to summarize the summary table output from the pipeline from multiple rows per specimen to a single row for each specimen, which is intended for lab staff to interpret and evaluate results for each specimen. The second R script looks up and appends specimen information from an Excel spreadsheet to the analysis pipeline summary table output, which is intended for use in evaluation and optimization of specimen preparation protocols.

REFERENCES
1. https://doi.org/10.1016/j.tim.2018.05.013
2. https://www.cdc.gov/flu/about/viruses/types.htm
3. https://www.nature.com/scitable/topicpage/genetics-of-the-influenza-virus-716/

Development/adoption of bioinformatics tools for pathogen surveillance systems.

AUTHORS
Andrew Lindsay[1], Stefan Gavriliuc[1], Xiaoli Dong[2], Vincent Li[1], Tarah Lynch[2] , Matthew Croxen[1]
[1] Alberta Precision Laboratories, Edmonton, Alberta, Canada
[2] Alberta Precision Laboratories, Calgary, Alberta, Canada

Presenting authors: andrew.lindsay@albertaprecisionlabs.ca , stefan.gavriliuc@albertaprecisionlabs.ca

KEYWORDS
Bioinformatics, pipeline, database, nextstrain, surveillance, automation

ABSTRACT
Epidemiological surveillance using genomics has emerged as a top priority for healthcare worldwide following the peak of the COVID-19 pandemic.[1] Hospitals and research laboratories are generating large amounts of sequence data, and there is increased demand for information systems that can parse and visualize such data in a rapid and automated fashion.

To meet this demand, we are working on an internal surveillance platform comprised of: bioinformatics pipelines for bacterial pathogens and metagenomics; a provincial sequencing database building on the work through the Genome Alberta 'Enabling Bioinformatics Solution' (EBS) project; and local instances of software like NextStrain[2] for rapid visualization and tracking of pathogen spread and evolution. This is showcased using 40k COVID samples from Alberta to visualize clade phylogeny, infection geography, sequence diversity and more.

The integration of these tools and approaches into a single platform will enable rapid, automated monitoring of a variety of pathogens within Alberta, with broad downstream applications for public health. Additional pipelines are being constructed to accommodate other pathogens such as Influenza or fungal sequence data. Future work will also include incorporating or developing further tools for pathogen genome analysis and visualization.

REFERENCES

1. Lo, S.W., Jamrozy, D. (2020). Genomics and epidemiological surveillance. *Nat Rev Microbiol* **18**, 478 (2020). https://doi.org/10.1038/s41579-020-0421-0
2. Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., Neher, R.A. (2018). Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**(23): 4121 – 4123. https://doi.org/10.1093/bioinformatics/bty407

Validation of a *Mycobacterium tuberculosis* bioinformatic pipeline for genomic surveillance.

AUTHORS
Colin Lloyd [1, 2, 3], Ashwin Deo [1], Silas Rotich[1, 3], Vincent Li[1], Sandra Shokoples[1], Cary Shandro[1], Gregory J. Tyrrell[1, 4], Matthew Croxen[1, 4].

[1] Alberta Precision Laboratories, Edmonton, Alberta, Canada
[2] Public Health Agency of Canada-National Microbiology Laboratory, Winnipeg, Manitoba, Canada
[3] University of Lethbridge, Lethbridge, Alberta, Canada
[4] Department of Laboratory Medicine and Pathology, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta, Canada

Presenting author: Colin.Lloyd@albertaprecisionlabs.ca

KEYWORDS
Mycobacterium tuberculosis, Surveillance, Antibiotic Resistance, cgMLST, Bioinformatic Pipeline

ABSTRACT
Introduction
Mycobacterium tuberculosis (MTB), the causative agent of tuberculosis, is a significant cause of morbidity and mortality worldwide and is a key public health pathogen. In addition, antimicrobial resistance to several frontline antimycobacterial antibiotics is a growing concern for the treatment of MTB. Currently MTB surveillance is done using phenotypic susceptibility testing and strain typing performed by analyzing specific genomic regions. Our study aims to modernize MTB surveillance through genome sequencing and validating a bioinformatic analysis pipeline to assess MTB antibiotic resistance markers and identify related isolates.

Methods
Nucleic acid was extracted from clinical MTB isolates cultured at Alberta Precision Laboratories – Provincial Laboratory North. Sequencing was performed using the Illumina DNA Prep kit and Illumina MiniSeq using a High Output flow cell. Sequences were analyzed using an in-house developed pipeline utilizing Kraken2, fastp, tb-profiler, seqkit, BWA, samtools, and kma-cgMLST to determine sequence read statistics, genome coverage and depth, presence of contamination, lineage assignment, resistance profiling, and clustering of related genomes. The H37Rv strain (NC_000962.3) was used as the reference genome.

Results
Sequences were accurately binned to MTB using Kraken 2 for each sample suggesting samples were free from contamination, and adequate sequencing was achieved with an average depth and genome coverage of >99% and >120X respectively. MTB genomes were assigned into 4 major lineages which were further characterized into 17 different sub lineages. The majority of genomes were categorized as antibiotic susceptible; however, genomes conferring both individual and multi-drug resistance were also identified. A total of 15 unique antibiotic resistance conferring mutations were identified. Comparing the genotypic resistance calls to phenotypic antibiotic susceptibility testing results, most samples matched, but discordance was also observed. A core-genome multi-locus sequence typing dendrogram was produced and accurately clustered genomes into distinct lineages.
Conclusion

Our MTB analysis pipeline was successful in identifying genomic antibiotic resistance markers and clustering of related isolates, which are both key components of MTB surveillance. However, additional validation is required to better understand the accordance between genotypic resistance markers and phenotypic antibiotic susceptibility as well as clustering of related isolates by inclusion of genomes from confirmed outbreaks.

Distinguishability of Episodic Fitness Shifts from Changes in Effective Population Size

AUTHORS
Oliver A.S. Lyon [1], A.P. Jason de Koning [1]

[1] Department of Biochemistry and Molecular Biology, University of Calgary, Calgary AB, Canada and Alberta Children's Hospital Research Institute (ACHRI), Calgary AB, Canada

Presenting author: oliver.lyon@ucalgary.ca

KEYWORDS
Phylogenetic inference, Effective population size, Population genetics, Fitness Shifts, Molecular evolution

ABSTRACT
The evolutionary process is influenced by many factors. However, not all forces of interest may be identifiable from comparative sequence data alone. Here we study what factors shape the distinguishability of changes in the underlying fitness landscape from variation in effective population size over time, since each of these processes can have similar influences on the expected distribution of sequence states at affected positions. Using a family of Markov modulated mutation-selection codon models with an explicit population genetic basis, we study distinguishability in terms of the Kullback-Leibler divergence between evolutionary models and extend this study using simulation. We thereby establish bounds on the number of sequence samples required before and after both fitness shifts and population size shifts on a large phylogenetic tree to achieve acceptable (best case) distinguishability and high power in resultant hypothesis tests. Our results highlight some of the challenges of modelling and inference of non-equilibrium molecular evolutionary processes from finite data.

A worm's view of genome assemblers with implications for gene identification and characterization.

AUTHORS
Grace M. Mariene[1], James D. Wasmuth[1]
[1] University of Calgary, Calgary, Canada
Presenting author: grace.mariene@ucalgary.ca

KEYWORDS
Nematode, heterozygosity, assemblers, annotation, genome

ABSTRACT
A genome assembly is an *in silico* approximation of the *in vivo* organization of an organism's DNA. The choice of best genome assembly software is contentious. In part, this is because performance varies between species, a consequence of different genome sizes, heterozygosity, and repeat content. The non-biological variability, aka error, introduced by assemblers has downstream effects on genome annotation and analysis, and can impact our ability to make biological predictions from the genome. With BioGenome projects using a one-size-fits-all approach to assemble genomes, we set out to quantify the variability across several long-read assemblers using three nematode species with different genome characteristics—size and complexity—and sequence reads generated on different platforms—PacBio and Oxford Nanopore.

We confirmed our prediction that no one assembler was 'best' for all species. We noticed that most assemblers gave similar scores when their assemblies were parsed through software which are supposed to determine accuracy. However, we compared the assemblies and found striking levels of variation in structure and number of genes contained within. One example is *Haemonchus contortus*, for which there is already a reference quality, chromosome-level assembly. However, we found that our less optimal assemblies contained highly conserved genes that were missing from the reference assembly. Across all three species, we found that in large multi-gene families, the number of members varied.

Through these results, we conclude that technical variation may look relatively minor but can have crucial consequences in the use of the genome data for subsequent studies. We urge researchers to adopt a 'let the buyer beware' principle to their use of genome assembly tsunami on its way.

A CURE for Introduction to Bioinformatics

AUTHORS

Tatiana Maroilley[1,2,3], Victoria Rodrigues Alves Barbosa[1,2,3], Diogo Marques[2], Catherine Diao[1,2,3], David Anderson[1,3*], Maja Tarailo-Graovac[1,2,3*]

[1] Department of Biochemistry and Molecular Biology, Cumming School of Medicine, University of Calgary, Calgary, Canada
[2] Department of Medical Genetics, University of Calgary, Calgary, Canada
[3] Alberta Children Hospital Research Institute, University of Calgary, Calgary, Canada
*co-last authors

Presenting author: tatiana.maroilley@ucalgary.ca

KEYWORDS
Transcriptome, Teaching, Structural variants, *C. elegans*, Undergrads

ABSTRACT
Bioinformatics is an interdisciplinary research area that combines Biology and Informatics[1]. Over the last decade, the advent of high-throughput technologies and big data has promoted the application of Bioinformatics methods to data of all research fields and of all scales – from molecules to populations[2]. Introducing Bioinformatics to undergraduate students in class is crucial, yet challenging as it requires students to acquire in a short time knowledge in coding, databases, technologies, biology, and biostatistics. Besides, only some students get a chance to practice Bioinformatics in research setting by joining laboratories over summer internships and Honor theses. However, such experiences are central for students' understanding of how Bioinformatics is applied in research.

In Winter 2023, we have implemented for the first time a Course-Based Undergraduate Research Experience (CURE) as an Introduction to Bioinformatics for second-year undergraduate students from the Bioinformatics program (Bachelor of Health Sciences (BHSc) at the Cumming School of Medicine at the University of Calgary) and beyond. In a class's setting, we offered the students a real research experience in Bioinformatics: from project design to dissemination, including advanced bioinformatics analyses of an original data set leading them to new discoveries.

Their projects consisted of exploring the impact of structural variants (SVs) such as translocation, deletion, or duplication on transcriptome data. In Maroilley et al.[3], we have previously used short-read whole genome sequencing to detect SVs in *Caenorhabditis elegans* (*C. elegans*) balancer strains. We then first trained the students in data manipulation, sequencing data analysis and interpretation to detect SVs by making them reproduce the Maroilley *et al*. [3] study. Then, they were asked to design an approach and analyze the transcriptome datasets of the same strains to scrutinize the potential impact of SVs on gene expression and splicing. With support of the teaching teams, they worked on this new RNA-Seq dataset and uncovered that heterozygous deletions do not always show a clear effect on gene expression level, as well as that translocation can lead to the activation of the expression of a new transcript. To complete the experience, we organized an in-person conference for students to present their findings to a large audience.

The first implementation of our CURE as an introduction to Bioinformatics revealed some challenges, especially technical (e.g., too large datasets for student personal laptops), but overall, it triggered students' engagement in class and gave them a unique opportunity to be part of a bioinformatics research project.

REFERENCES

1. https://pubmed.ncbi.nlm.nih.gov/11552348/
2. https://pubmed.ncbi.nlm.nih.gov/33735179/
3. https://pubmed.ncbi.nlm.nih.gov/34521941/

Differential adenosine to inosine RNA editing of SINE B2 non-coding RNAs in mouse unveils a novel type of epi-transcriptome response to amyloid beta neuro-toxicity

AUTHORS
Liam Mitchell[123], Luke Saville[123], Yubo Cheng[123], Travis Haight[123], Riya Roy[1234], Cody Turner[123], Gerlinde A.S. Metz[24], Majid Mohajerani[4], Athanasios Zovoilis[1234]

[1] Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, Alberta, Canada
[2] Southern Alberta Genome Sciences Centre, Lethbridge, Alberta, Canada
[3] Canadian Centre for Behavioral Neuroscience, Department of Neuroscience, Lethbridge, Alberta, Canada
[4] Alberta RNA Research and Training Institute, Lethbridge, Alberta, Canada

Presenting author: liam.mitchell@uleth.ca

KEYWORDS
RNA modification, epi-transcriptomics, inosine, repetitive elements, bioinformatics

ABSTRACT
Alzheimer's disease (AD) is neurodegenerative disorder, whose underlying molecular pathology is largely unclear. The study of standard-dogma molecular biology (DNA→RNA→Protein) has largely been unable to resolve the underlying mechanisms of such complex multi-factor diseases. Recent advancements in sequencing technology have made it possible to study other kinds of molecular control mechanisms, such as RNA editing. The type of RNA editing involving conversion from adenosine to inosine (A-to-I) has been connected with cellular function. Specifically in domains of cell-mediated immunity, dsRNA homeostasis, and cellular stress. Among the RNAs that have been reported to harbor a large percentage of A-to-I edits are non-coding RNAs generated from Short Interspersed Nuclear Elements (SINEs), such as B2 RNAs in mouse and Alu RNAs in human. We have recently shown that B2 RNAs can act as riboswitches, regulating gene expression though self-cleavage, and they are abnormally processed during amyloid beta pathology in hippocampal cells, contributing to the transcriptome de-regulation observed in this condition. Here, we present a novel A-to-I editing analysis approach, which is customized for repetitive elements, and we report that changes in A-to-I editing of SINE B2 RNAs is associated with epi-transcriptome response to amyloid beta neurotoxicity and pathology. We show that differential rates of A-to-I editing in B2 RNA is increased as an early response to amyloid neural toxicity in both the hippocampi of a mouse model of amyloid beta pathology and a hippocampal cell culture model of amyloid beta toxicity. This data suggests that the recently described mode of gene-expression-regulation through B2 RNA processing may be intertwined with RNA editing, and that the cell may be employing RNA editing as protective mechanism to counteract the effect of increased B2 RNA processing during the initial phases of response to amyloid beta toxicity. To this end, inhibiting A-to-I editing activity in hippocampal cells results in increased destabilization of B2 RNAs. Our findings unveil RNA editing of SINE RNAs as an additional level of epi-transcriptome response to amyloid beta neuro pathology, with potential implications for the role of RNA editing of SINE RNAs also in human and Alzheimer's disease (AD).

Comparative genomics reveals the widespread nature of an unappreciated membrane trafficking pathway in protists

AUTHORS

Kiran J. More[1], Joel B. Dacks[1], Paul T. Manna[1,2]
[1] University of Alberta, Edmonton, Canada
[2] University of Gothenburg, Gothenburg, Sweden

Presenting author: kiran.more@ualberta.ca

KEYWORDS

Membrane trafficking, similarity searches, cell biology, evolution, lysosome-related organelles

ABSTRACT

A complex endomembrane system is one of the hallmarks of eukaryotic cells. In most eukaryotes (the vast majority of which are unicellular, microscopic protists), the endomembrane system includes a conserved suite of organelles (e.g., the Golgi apparatus, endoplasmic reticulum, lysosome). Other organelles, such as the lysosome-related organelles (LROs), can confer unique functions in specific cells or tissues. LROs are heterogenous in biogenesis and function but have the common feature of deriving most of their components through the endolysosomal system1. Outside of the few animal-specific model LROs, the knowledge of molecular machinery associated with these lineage-specific organelles is sparse.

The Biogenesis of Lysosome-related Organelle Complexes (BLOCs) are a set of protein complexes that are united through causing the human genetic disease Hermansky-Pulak Syndrome1, which results in disordered trafficking of proteins normally targeted to certain animal-specific LROs. The BLOCs have been best characterized and largely found in animals, but a few of the components have previously been reported in other lineages2, raising the possibility that these complexes are more widely distributed. Using comparative genomics methods, we investigated the distribution of the BLOCs using an updated and comprehensive selection of 139 eukaryotic genomes and transcriptomes. Though known functions of these complexes are restricted to animal-specific LROs, we show that BLOCs are widely distributed—but not widely conserved—across the broader diversity of eukaryotes. Sensitive similarity searching, domain structure inference, and a review of known BLOC interactors suggest homology between BLOC-2 and HOPS/CORVET, and their divergence prior to the diversification of extant eukaryotes. The function of the BLOC trafficking pathway remains completely uncharacterized outside of animal models, which opens an intriguing new area of research to further our understanding of protist LROs.

REFERENCES

1. https://doi.org/10.1111/tra.12646
2. https://doi.org/10.1111/j.1600-0854.2010.01044.x

Cooperative Interactions of Glucocorticoid Receptor and NF-κB at Overlapping Genomic Loci: Insights from ChIP-seq Analysis

AUTHORS
Mahmoud M. Mostafa, Andrew Thorne, Akanksha Bansal, Robert Newton
University of Calgary, Calgary AB, Canada
Presenting author: mahmoud.mostafa@ucalgary.ca

KEYWORDS
glucocorticoids, inflammation, NF-κB, transcription factor binding, ChIP-seq

ABSTRACT
INTRODUCTION: Glucocorticoids (GCs) are hormones essential for normal physiology and, clinically, they serve as potent anti-inflammatory drugs. They exert their effects by binding to the glucocorticoid receptor (GR) which, upon ligand activation, translocates to the nucleus and binds to thousands of genomic loci. This leads to transcriptional activation of hundreds of genes, an effect that found to be crucial in reversing most of the transcriptional effects of pro-inflammatory transcription factors like NF-κB and AP1. NF-κB, a family of transcription factors, regulates genes involved in immune response, inflammation, and cell survival. Activated by various stimuli, including cytokines, NF-κB heterodimers (e.g., p65 (RELA) and p50) translocate to the nucleus and bind specific DNA sequences to induce the expression of target genes. When both factors are activated, GR and RELA were reported to occupy overlapping genomic loci, and their interactions have been hypothesized to result in transcriptional repression of NF-κB-dependent genes[1]. However, as such hypothesis fails to explain most of the anti-inflammatory effects of glucocorticoids[2], the true nature of these interactions remains unclear.

RESULTS: Using ChIP-seq, we assessed genome-wide binding of GR and RELA following 1h treatment with IL1B (activating RELA), budesonide (synthetic GC), or a combination of both. Our analysis revealed that consensus GR and RELA binding motifs were highly enriched in their respective binding sites following either mono- or combo- treatment, confirming the validity of our results and suggesting no major shift in binding behavior due to combination treatment. While ~25% of RELA sites (following IL1B) separately recruited GR (following budesonide), the number of common sites nearly tripled after combination treatment. Moreover, combination treatment increased GR and RELA recruitment predominantly at overlapping sites, while decreasing binding at non-overlapping sites. The change in GR or RELA recruitment following combination treatment was positively correlated with the binding intensity of the other factor, with GR binding motifs being dominant in increasing RELA sites and vice versa.

CONCLUSIONS: Our findings suggest that GR and RELA interactions are cooperative rather than antagonistic, as both factors facilitate the recruitment of the other factor to their binding sites, and both are considered to mainly act as transcriptional activators. Further investigation of transcriptional activation using RNAP2 ChIP-seq or GRO-seq, and gene expression analysis with RNA-seq, is needed to confirm the cooperative nature of these interactions and to better understand their functional consequences. This insight may have important implications for better management of GC-resistant inflammation, such as severe asthma.

REFERENCES

1. https://genome.cshlp.org/content/21/9/1404
2. https://doi.org/10.1016/j.jbc.2021.100687

Jumping jellyfish: *Myxobolus rasmusseni*, a cnidarian parasite of Albertan fish, has a genome that is highly repetitive, full of transposons, and has experienced massive gene loss

AUTHORS

Viraj Muthye[1], Aralia Leon Coria [1], Hongrui Liu [1], Molly Tilley [2], Constance A. M. Finney [1], Cameron P. Goater [2], James D. Wasmuth [1]

[1] University of Calgary, Calgary, Alberta, Canada
[2] University of Lethbridge, Lethbridge, Alberta, Canada

Presenting author: virajrajendra.muthye@ucalgary.ca

KEYWORDS

myxozoa, parasites, genome, repeats, nanopore

ABSTRACT

Myxozoans are an enigmatic group of parasites within Cnidaria, a phylum that also includes jellyfish, corals, and sea anemones. In their life cycle, myxozoans alternate between invertebrates (annelids or bryozoans) and vertebrates (including several economically important fish species). For example, *Myxobolus cerebralis*, which causes Whirling Disease in salmonid fish, inflicts mortality rates of up to 90% in juveniles, leading to severe economic losses in aquaculture, risking food security, and closing rivers to recreational fishing. Currently, there are no effective treatments against myxozoan infections. Despite this, the genomes of only eight of the ~2,400 myxozoan species have been sequenced. Here, we have used Oxford Nanopore MinION to sequence the genome of *Myxobolus rasmusseni* n sp., an emerging myxozoan parasite of fathead minnows which was recently discovered in southern Alberta. The assembled *M. rasmusseni* genome is 171 Mb, with N50 of 422.2 Kb, and one of the largest myxozoan genomes reported. In our annotation of the genome, we identified 9,529 protein-coding genes and found that, like other myxozoan genomes, *M. rasmusseni* has lost about 45% of the most highly conserved eukaryote genes. Surprisingly, nearly 69% of the genome was comprised of repeat elements, with 43% of the genome derived from DNA transposons, making it one of the most repetitive genomes in animals. We also reannotated the genomes of the other myxozoan species and identified considerable genome plasticity within this taxonomic group. We are now focusing on using these genomes to find new treatment strategies that are pan-myxozoan.

Semi-Self-Supervised Learning for Semantic Segmentation in Images with Dense Patterns

AUTHORS

Keyhan Najafian[1], Farhad Maleki[2], Ian Stavness[1], Lingling Jin[1*]
[1] University of Saskatchewan, Saskatoon, Canada
[2] University of Calgary, Calgary, Canada

Presenting author: keyhan.najafian@usask.ca

KEYWORDS

Semantic Segmentations, Semi-Supervised Learning, Self-Supervised Learning, Dense Pattern, Wheat Head Segmentation, Domain Adaptation, Deep Learning

ABSTRACT

Deep learning has shown potential in domains with large-scale annotated datasets. However, manual annotation is expensive, time-consuming, and tedious. Pixel-level annotations are particularly costly for semantic segmentation in images with dense irregular patterns of object instances, such as in plant images. In our paper "Semi-Self-Supervised Learning for Semantic Segmentation in Images with Dense Patterns" recently published on Plant Phenomics [1], we proposed a method for developing high-performing deep learning models for semantic segmentation of such images utilizing little manual annotation. The application of the method focused on the wheat head segmentation task as a use case. We synthesized a computationally-annotated dataset---using only a few annotated images, a short unannotated video clip of a wheat field, and several video clips with no wheat---to train a customized U-Net model. Considering the distribution shift between the synthesized and real images, we applied three domain adaptation steps to gradually bridge the domain gap. As the first domain adaptation step, we developed a new dataset of real images of wheat plants, showcasing their complete structure and positional attributes, along with suitable transformations, including a rotation at angles from 0 to 360 degrees, and applying online augmentation transformations. We used new data to fine-tune the model trained on synthesized data. In the second step, the trained model was used for pseudo-labeling and predicting segmentation masks for the image frames extracted from our only wheat field video. These pseudo-labeled samples were subsequently used for fine-tuning the model. In the final stage, two samples from each domain of 18 distinct external domains were utilized to further address the domain shift between synthesized and real images. These resulting sets were employed for fine-tuning the previously trained model. Only using two annotated images, we achieved a Dice score of **0.89** on the internal test set. When evaluated on a diverse external dataset collected from 18 different domains across five countries, this model achieved a Dice score of **0.73**. To expose the model to images from different growth stages and environmental conditions, we incorporated two annotated images from each of the **18 domains** to further fine-tune the model. This increased the Dice score to **0.91**. The result highlights the utility of the proposed approach in the absence of large-annotated datasets. Although our use case is wheat head segmentation, the proposed approach can be extended to other segmentation tasks with similar characteristics of irregularly repeating patterns of object instances.

REFERENCES

https://doi.org/10.34133/plantphenomics.0025

SLiM-Tree: Simulating Molecular Evolution along Phylogenies with Pure Population Genetics Models and Realistic Fitness Functions

AUTHORS
Afarinesh Panahy*, Erin Brintnell*, Grace Heemeryck, Dave Anderson, and AP Jason de Koning.
Department of Biochemistry and Molecular Biology, Cumming School of Medicine, University of Calgary, Alberta, Canada.

Presenting author: afarinesh.panahy@ucalgary.ca

KEYWORDS
PopulaKon GeneKcs, EvoluKon, MutaKon rate, Theta, SLiM-Tree

ABSTRACT
We present a Python package for automating the use of SLiM to conduct pure population genetics simulations along phylogenies, SLiM-Tree. One application for this work is to enable examination of the realism of the many simplifying assumptions embedded in all models of molecular evolution by phylogenetic substitution. In my thesis project, I am examining whether there is systematic bias in standard estimators of population mutation rates in populations where is large. We are particularly interested in verifying whether some natural populations have large enough that the mean time between mutations is close to, or shorter than, the time it takes for mutations to go to their fates. Theoretical work from our lab has controversially shown that when , the true rate of substitution slows down relative to Kimura's infinite sites rate of substitution (or its weak mutation finitesites generalization), making the rate of neutral evolution less than the mutation rate [1]. In this regime, the rate of non-neutral evolution for adaptive alleles slows down ever more relative to weak mutation/ infinite sites theory, which may introduce bias in a variety of standard methods for inferring the strength of natural selection in the genome in populations with large , such as in mutation-selection codon models and with standard type approaches.

My work will examine what evidence there is for real populations with , and will utilize a mix of pure population genetics simulations of molecular evolution (using our program, SLiM-Tree), and inference of population mutation rates with a variety of widely used estimators such as Watterson's, Nei and Tajima's and those implemented in BEAST.

REFERENCES
1. https://www.biorxiv.org/content/10.1101/259507v2

Biomarker discovery workflow for chemical isotope labeling LC-MS metabolomics

AUTHORS
Rui (Ric) Qin[1], Wayne Cheng[1], Shuang Zhao[1], Liang Li[1, 2]

[1] The Metabolomics Innovation Centre (TMIC), Edmonton, Canada
[2] University of Alberta, Edmonton, Canada

Presenting author: rq2@ualberta.ca

KEYWORDS
Untargeted metabolomics, chemical isotope labelling, biomarker discovery, feature selection, ensemble model, machine learning

ABSTRACT
Untargeted metabolomics, aiming to characterize and quantitate a broad spectrum of metabolites in biological samples, has become an important tool in the discovery of biomarkers for environmental exposures and diseases[1]. The chemical isotope labeling (CIL) LC-MS method developed by our group significantly improves the sensitivity and metabolite coverage in untargeted metabolomics[2], enhancing our potential in identifying genuine biomarkers. However, high numbers of metabolites identified by CIL LC-MS present a challenge for selection of relevant features. Compared to conventional feature selection methods, ensemble feature selection approaches can produce more stable and accurate results[3], and are only recently applied to untargeted metabolomics[4] as well as other omics studies. Here, we present an ensemble model-based biomarker discovery workflow integrated with our CIL-MS pipeline. Tested on serum CIL LC-MS data of several independent rheumatoid arthritis (RA) cohorts, this ensemble feature selection approach showed high robustness and high versatility in generating panels of biomarker candidates based on different statistical metrics.

REFERENCES

1. https://doi.org/10.3389/fbioe.2015.00023
2. https://doi.org/10.1021/acs.analchem.9b03431
3. https://doi.org/10.1186/s12859-022-05132-9
4. https://doi.org/10.1109%2FISSPIT.2017.8388679

Genomic investigation of increased *Streptococcus pyogenes* infection in Alberta.

AUTHORS
Silas K. Rotich[1,4], Petya Koleva[1,2], Vincent Li[1], Ashwin Deo[1], Colin Lloyd[1,2,4], Sandy Shokopoles[1], Gregory J. Tyrrell[1,3], Mathew Croxen[1,3].
[1] Alberta Precision Laboratories, Edmonton, Canada
[2] Public Health Agency of Canada – National Microbiology Laboratory, Winnipeg, Canada
[3] Department of Laboratory Medicine and Pathology, University of Alberta, Edmonton, Canada
[4] University of Lethbridge, Lethbridge, Canada

Presenting author: Silas.Rotich@albertaprecisionlabs.ca

KEYWORDS

Streptococcus pyogenes, Group A Streptococcus in Alberta, M1UK, exotoxins, Surveillance.

ABSTRACT

INTRODUCTION
Streptococcus pyogenes, commonly referred to as Group A Streptococcus (GAS), is a bacterium that can cause a range of infections from mild, such as impetigo and strep throat, to severe, such as necrotizing fasciitis and toxic shock syndrome. In recent times, GAS has become a bacterium of concern due to emergence of a new variant of *S. pyogenes* M1 (designated '$M1_{UK}$') that has been identified in the United Kingdom, particularly in pediatric populations. The $M1_{UK}$ strain has been linked to increased seasonal scarlet fever and invasive infections, marked by high expression of the superantigen s*peA*, a critical virulence factor. The $M1_{UK}$ strain has been reported to contain specific mutations that could potentially distinguish it from other M1 strains. In this study, we investigated confirmed GAS isolates in Alberta to determine whether they exhibit distinctive features of the $M1_{UK}$ strain.

METHODS
Clinical GAS (n=55) isolates were prospectively cultured at Alberta Precision Laboratories – Provincial Laboratory North and nucleic acid extracted as part of an on-going surveillance. Sequencing was performed on the Illumina MiniSeq using the Illumina DNA Prep kit with a rapid protocol adapted for bacterial sequencing. To characterize GAS isolates, sequences were subject to ad hoc. analyses in order to classify the isolates as $M1_{UK}$ or not based on 27 defined mutations in the suspect genomes. Phylogenies were then constructed using core single nucleotide variants, and the presence of 11 exotoxins (s*meZ, speA, speB, speC, speG, speH, speI, speJ, speL, speR* and *ssa*) was assessed. *De novo* assembly of each genome, and *in silico* emm-typing and MLST was also performed.

RESULTS
Analysis of the sequences identified the $M1_{UK}$ strain among M1 global based on the presence of 27 specific mutations. Phylogenetic analysis based on core single nucleotide variants confirmed that these $M1_{UK}$ isolates clustered together and were distinct from other M1 strains. *De novo* assembly, *in silico* emm-typing and MLST further confirmed that these isolates were serotype emm1 and ST28. s*peA, speB, speG* and *speI* were detected in both $M1_{UK}$ and M1 global strains, but *speC* was more commonly found in $M1_{UK}$ than M1 global.

CONCLUSION

Overall, these results suggest that the $M1_{UK}$ strain is present in Alberta and maybe associated with increased virulence. The identification and characterization of this strain in Alberta highlights the importance of ongoing surveillance for GAS infections and underscores the need for continued efforts to monitor the emergence and spread of virulent GAS strains.

Elucidating the role of Adenosine-to-Inosine editing in SINE RNAs: Connecting the dots between editing and cellular response to stress

AUTHORS:
Riya Roy (1,2 ,3), Travis Haight (1,2 ,3), Luke Saville (1,2 ,3), Liam Mitchell (1,2 ,3), Athanasios Zovoilis

Affiliations:
(1) Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, AB, Canada
(2) Southern Alberta Genome Sciences Centre, University of Lethbridge, Lethbridge, AB, Canada
(3) Canadian Centre for Behavioral Neuroscience, University of Lethbridge, Lethbridge, AB, Canada

Presenting author: riya.roy@uleth.ca

KEYWORDS:
Non-coding RNAs, SINEs, RNA editing, Murine B2, Cellular stress

ABSTRACT:
Short Interspersed Nuclear Elements (SINEs) are one of the most abundant classes of non-coding RNAs involving human Alu and murine B2 as the most frequent ones[1]. Adenosine-to-Inosine RNA editing by adenosine deaminase acting on RNA (ADAR) enzymes, is a very common post-transcriptional modification that predominantly targets double stranded RNA that majorly includes Alu and B2 SINEs[3]. Recent studies have revealed that SINE RNAs play a key role in cellular response to stress[2], however the role of A-to-I editing in SINE RNA stability remains unclear. Here, we are trying to understand the connection between RNA editing and the stability of B2 SINE RNAs, focusing more on the correlation between SINE RNA processing ratio and A-to-I editing during response to cellular stress in mouse fibroblasts.

REFERENCES:

1. Yakovchuk P, Goodrich JA, Kugel JF. B2 RNA and Alu RNA repress transcription by disrupting contacts between RNA polymerase II and promoter DNA within assembled complexes. Proc Natl Acad Sci U S A. 2009 Apr 7;106(14):5569-74. doi: 10.1073/pnas.0810738106. Epub 2009 Mar 23. PMID: 19307572; PMCID: PMC2667051.
2. Cheng Y, Saville L, Gollen B, Isaac C, Belay A, Mehla J, Patel K, Thakor N, Mohajerani MH, Zovoilis A. Increased processing of SINE B2 ncRNAs unveils a novel type of transcriptome deregulation in amyloid beta neuropathology. Elife. 2020 Nov 16;9:e61265. doi: 10.7554/eLife.61265. PMID: 33191914; PMCID: PMC7717908.
3. Huntley MA, Lou M, Goldstein LD, Lawrence M, Dijkgraaf GJ, Kaminker JS, Gentleman R. Complex regulation of ADAR-mediated RNA-editing across tissues. BMC Genomics. 2016 Jan 15;17:61. doi: 10.1186/s12864-015-2291-9. PMID: 26768488; PMCID: PMC4714477.

Solving the genomics of unsolved rare life-threatening COVID-19 using genome sequencing

AUTHORS
Li Shu[1,2], Jamie Leckie[3], Tatiana Maroilley[1,2], Gerald Pfeffer[2-4], Maja Tarailo-Graovac[1,2]
[1]Departments of Biochemistry, Molecular Biology and Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, AB T2N 4N1, Canada
[2]Alberta Children's Hospital Research Institute, University of Calgary, Calgary, AB T2N 4N1, Canada
[3] Hotchkiss Brain Institute, University of Calgary, Calgary, AB, T2N 4N1
[4] Department of Clinical Neurosciences, and Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary AB, T2N 4N1

Presenting author: li.shu@ucalgary.ca

KEYWORDS
genome sequencing, COVID-19, genetics, life-threatening, rare disease

ABSTRACT
Background: COVID-19, caused by the SARS-CoV-2 virus, has been declared a pandemic for several years and has been considered a global health crisis. The phenotypes of COVID-19 are highly heterogeneous, posing a need to delimit potential genetic predisposition to severe COVID-19 outcomes for early recognition and timely clinical intervention. We hypothesize that severe illness with COVID-19 in some cases which do not otherwise have pre-existing risk factors is attributable to monogenic causes.

Methods: Research participants were from the Alberta Host Genetic Susceptibility to COVID-19 project (AB-HGS), a participating study in Hostseq[1]. Prospective participants were identified using administrative health data meeting certain criteria (age<65 years, required admission to hospital for COVID-19, and absence of known medical comorbidities) and contacted by post. Participants contacted researchers to participate in the study and consented to provide clinical information and a DNA sample for whole genome sequencing (WGS). Family members with mild illness were recruited as controls.

Results: A total of 73 participants were included in the study. For our initial analysis of WGS data, we identified a panel of 455 genes associated with COVID-19 outcome after searching published articles in PubMed using keywords "genetic" and "severe COVID-19" inclusive of April 19th, 2023. We have successfully applied WGS pipeline and analyzed the rare coding variants from the 455 gene panel. We identified variants in *SPEG, TLR7, NADSYN1* genes which were shared in two family members in the same pedigree. The two members were composed of one hospitalized severe patient who carried a single rare variant and one non-hospitalized person who carried not only the same rare variant but the other variant from the panel genes (*SPEG/GSTT1; TLR7/STAT2; NADSYN1/FGB*). We identified a rare *CHD5* variant in two severe patients and one non-hospitalized person who also carried another variant in *LAMB1* and *GOLM1*. We also identified *DNAH7* homozygous variants and *ASH1L* compound heterozygous variants in two severe patients separately.

Discussion: Further validation is in progress as well as extended analysis to include structural and non-coding variants. Correlation to clinical data may yield additional insights. The approach may be applied broadly to other data within HostSeq, which includes nearly 10,000 participants. As such, our work has the potential to build a diagnostic workflow for rare, life-threatening COVID-19 patients and provide clues for identifying individuals predisposed to severe COVID-19 outcomes.

REFERENCES

1. https://pubmed.ncbi.nlm.nih.gov/37131148/

Identifying Key Genes in Preeclampsia using Machine Learning

Anureet Tiwana

University of Alberta, Edmonton, Canada
atiwana@ualberta.ca

KEYWORDS

Preeclampsia, Random Forest, Support Vector Machine, Generalized Linear Models, and Key Genes

ABSTRACT

Preeclampsia is a dangerous condition that affects pregnant women after 20 weeks of gestation, marked by proteinuria and high blood pressure. The disease can lead to various problems such as early birth, HELLP syndrome, kidney damage, and eclampsia [1-2]. Low vitamin D status and other risk factors are associated with the development of preeclampsia. The condition affects 5-8% of pregnancies in the United States and 8 million pregnancies worldwide, leading to 60,000 maternal deaths per year [3]. Preeclampsia affects fetal development and growth, frequently resulting in perinatal and neonatal morbidity or mortality and contributing to preterm births [4]. The disease has significant effects on brain development and functions of offspring over time, resulting in intellectual disability, epilepsy, autism, and schizophrenia [5]. The molecular mechanisms of preeclampsia in the placenta are still mainly unknown, and there are insufficient molecular criteria for distinguishing clinical subtypes [6]. The placental injury caused by hypoxia/reperfusion is consistent with a pathological inflammatory response, which can cause a systemic inflammatory response and endothelial damage, contributing to the development of preeclampsia. Identifying differentially expressed genes and enhancing their biological activities and critical pathways will be the first step toward understanding the molecular mechanisms of preeclampsia. A study aims to identify key genes from datasets using differential gene expression analysis and machine learning models such as Support Vector Machine, Random Forest, and Generalized Linear Models. The study design involves data preprocessing, differential gene expression analysis, machine learning models, result comparison, and visualization. The visualization step involves visualizing the expression patterns of identified key genes and creating plots to analyze their biological significance using functional enrichment analysis.

REFERENCES

1. Rana S., Lemoine E., Granger J. P., Karumanchi S. A. Preeclampsia: pathophysiology, challenges, and perspectives. *Circulation Research* . 2019;**124**(7):1094–1112. doi: 10.1161/CIRCRESAHA.118.313276.
2. Phipps E. A., Thadhani R., Benzing T., Karumanchi S. A. Pre-eclampsia: pathogenesis, novel diagnostics and therapies. *Nature Reviews Nephrology* . 2019;**15**(5):275–289. doi: 10.1038/s41581-019-0119-6.
3. Weinstein L. Syndrome of hemolysis, elevated liver enzymes, and low platelet count: a severe consequence of hypertension in pregnancy. Am J Obstet Gynecol. 1982 Jan 15;142(2):159-67. doi: 10.1016/s0002-9378(16)32330-4.
4. Wallis AB, Saftlas AF, Hsia J, Atrash HK. Secular trends in the rates of preeclampsia, eclampsia, and gestational hypertension, United States, 1987-2004. *Am J Hypertens.* 2008;**21**(5):521–526. doi: 10.1038/ajh.2008.20.
5. Jeyabalan A. Epidemiology of preeclampsia: impact of obesity. *Nutr Rev.* 2013;**71**(Suppl 1):S18–S25.
6. Shah DA, Khalil RA. Bioactive factors in uteroplacental and systemic circulation link placental ischemia to generalized vascular dysfunction in hypertensive pregnancy and preeclampsia. *Biochem Pharmacol* 95: 211– 226, 2015. doi:10.1016/j.bcp.2015.04.012.

SV-JIM: a comprehensive structural variant detection pipeline to reduce variance between multiple SV callers.

AUTHORS

C. Malcolm Todd[1], Lingling Jin[1*]
[1] Department of Computer Science, University of Saskatchewan, Saskatoon, SK, S7N 5C9, Canada

Presenting author: C. Malcolm Todd - gwy260@mail.usask.ca

KEYWORDS
Structural variants, genetic variation, comparative genomics, plant genomes, sequence alignment

ABSTRACT
Structural variations (SVs) are large-scale sequence rearrangements that play a significant role in the diversity and evolution due to their ability to affect a species' functional gene content. In fact, SVs have been linked in multiple studies to many important biological traits and serious diseases. For example, the presence of SVs has been linked to diabetes [1], cancers [2], and autism spectrum disorders [3] in humans, but also climate adaptability and yield in plants [4]. Despite their importance, however, there remains significant variance in the outputs produced by SV identification tools due to the complex nature of the SV identification task and there remains no single best tool capable of identifying SVs in all situations. Furthermore, most SV tools were developed and tested based on the human genome; however, SV identification remains a challenging area in plant genomes that requires further study due to the highly repetitive and polyploidy nature of plant genomes that can reduce SV tool accuracy. To that end, we propose a new pipeline named SV-JIM that combines evidence from multiple read- and assembly-based SV tools including cuteSV [5], PAV [6], Sniffles2 [7], SVIM [8], and SVIM-asm [9] to help eliminate result variance and to identify locations of SVs supported by multiple tools. SV-JIM also streamlines the execution and configuration of the SV calling process by implementing the pipeline using the Snakemake [10] workflow management system. During execution, SV-JIM produces a series of SV sets that provides users with greater flexibility in picking the preferred combination or aggregation of SV tools to increase the probability that the generated results meet their needs. During this work, we also executed SV-JIM on two plant genomes including *Brassica nigra* and *Arabidopsis thaliana* to validate the pipeline's execution and to assess the amount of variance between the tools when applied to real genome data. These results served as a valuable example of the high variance between SV tools with results ranging between 26,516 and 96,322 SVs for *B. nigra* and between 3,514 and 8,713 SVs for *A. thaliana*. In addition, these results yielded further insights that helped to identify certain differences in reporting and configuration options between the included tools that may contribute to this variance in their results and allowed SV-JIM to identify which combinations of tools were most consistent or most divergent compared to the others. SV-JIM is available publicly under the MIT license on GitHub at https://github.com/CMalcolmTodd/SV-JIM.

REFERENCES

1. Long-read sequencing of 3,622 Icelanders provides insight into the role of structural variants in human diseases and other traits. (Beyter et al.) (https://doi.org/10.1038/s41588-021-00865-4).
2. Sequencing structural variants in cancer for precision therapeutics. (Macintyre et al.) (https://doi.org/10.1016/j.tig.2016.07.002)

3. Integrated functional analysis implicates syndromic and rare copy number variation genes as prominent molecular players in pathogenesis of autism spectrum disorders. (Ashitha et al.) (https://doi.org/10.1016/j.neuroscience.2020.04.051)
4. Exploring and exploiting pan-genomics for crop improvement. (Yongfu et al.) (https://doi.org/10.1016/j.molp.2018.12.016)
5. Long-read-based human genomic structural variation detection with cuteSV. (Jiang et al.) (https://doi.org/10.1186/s13059-020-02107-y)
6. Haplotype-Resolved Diverse Human Genomes and Integrated Analysis of Structural Variation. (Ebert et al.) (https://doi.org/10.1126/science.abf7117)
7. Accurate detection of complex structural variations using single-molecule sequencing. (Sedlazeck et al.) (https://doi.org/10.1038/s41592-018-0001-7)
8. SVIM: structural variant identification using mapped long reads (Heller et al.) (https://doi.org/10.1093/bioinformatics/btz041)
9. SVIM-asm: structural variant detection from haploid and diploid genome assemblies (Heller et al.) (https://doi.org/10.1093/bioinformatics/btaa1034)
10. Sustainable data analysis with Snakemake (Mölder et al.) (https://doi.org/10.12688/f1000research.29032.2)

mRNA expression profiling and functional enrichment analysis of cattle liver abscess

AUTHORS
Yixin Wang[1,2], Jian Wang[3], Zhe Pan[1], Robert J. Gruninger[4], Rahat Zaheer[4], Tim A. McAllister[4], Mingzhou Li[2], Le Luo Guan[1*]
[1]Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Alberta, T6G 2P5, Canada;
[2]Institute of Animal Genetics and Breeding, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, Sichuan, 611130, China;
[3]State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, College of Animal Science and Technology, Guangxi University, Nanning, Guangxi, 530004, China;
[4]Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, Lethbridge, Alberta, T1J 4B1, Canada

Presenting author: yixin25@ualberta.ca

KEYWORDS
Cattle liver abscess, RNA-seq, mRNAs, differential expression analysis, functional enrichment

ABSTRACT
The molecular mechanism of cattle liver abscess pathogenesis is still unknown, the prevalence of which exceeds 25% and causes economic losses to the domestic industry. Therefore, to detect the potential molecular factors related to cattle liver abscesses, 16 liver samples were collected from healthy beef cattle (HH), 16 were collected from the abscesses region of abscessed livers from cattle (AA), besides 16 were collected from non-abscessed regions of these abscessed livers (AH). Totally, 48 rRNA-depleted libraries were constructed for RNA sequencing (PE100). After quality control, about 35 million clean reads were obtained from each sample and aligned to the cattle reference genome (ARS-UCD1.2). Nearly 90% of clean reads were uniquely mapped and then used for cattle mRNA quantification. A total of 12,666 of 21,861 mRNA genes were defined as expressed genes under the threshold that the transcripts per million (TPM) > 1 and detected at least in 80% of all samples. The hierarchical cluster analysis showed a similar expression pattern between AH and HH groups, while the samples in the AA group were separated from AH and HH cluster. Moreover, we used edgeR and set a strict threshold (The expressed genes with False Discovery Rate (FDR) < 0.01 and log2 fold change > |2|) for differentially expressed genes (DE-genes) detection. The comparison between AA and HH groups identified 1,630 DE-genes, with 1,151 DE-genes upregulated and 479 DE-genes downregulated in AA. The comparison between AA and AH groups identified 1,442 DE-genes, with 1,103 DE-genes upregulated and 339 DE-genes downregulated in AA. However, no DE gene was identified between AH and HH. Functional enrichment analysis was performed by DAVID, and upregulated DE-genes in AA compared with HH group were involved in inflammation, the extracellular matrix, and immune response, while downregulated DE-genes in AA were involved in metabolism terms like glucuronosyltransferase activity and retinol metabolism. The enrichment function of DE-genes between AA and AH was similar to those between AA and HH. The differences in mRNA expression profile indicate impaired metabolism function of the abscess region of the liver, which undergoes inflammation, while non-abscessed regions maintained similar mRNA expression profiles and normal function with healthy liver tissue.

Comparative and functional genomics analysis revealed host specific carbohydrate metabolism of bovine-origin *Bifidobacterium longum* subsp. *longum* strains.

Anna Widenmann, Mi Zhou, Michael Gänzle, Le Luo Guan
Department of Agricultural, Food and Nutritional Sciences, University of Alberta, Edmonton, Canada

Presenting author: Anna Widenmann

KEYWORDS

Genomics, Bovine, Carbohydrate active enzymes, dbCAN, Carbohydrate metabolism, Comparative genomic analysis, Probiotics

ABSTRACT

The two subspecies of *Bifidobacterium longum*, *Bifidobacterium longum* subsp. *longum* and *Bifidobacterium longum* subsp. *infantis* are pioneer colonizers of the animal gut [1]. A large portion of their functional genes are attributed to the breakdown and uptake of carbohydrates; however, little is known about the carbohydrate active enzymes (CAZymes) involved in the carbohydrate metabolism of bovine *Bifidobacterium* strains [2]. Therefore, the objective of this work was to describe diverse functional capacities in carbohydrate metabolism of different bovine-derived *Bifidobacterium longum* subsp. *longum* strains by comparative genomics.

Four *Bifidobacterium longum* subsp. *longum* strains were isolated from the colon of healthy calves and cultivated anaerobically in mMRS with 0.05% L-cysteine·HCl at 37°C. DNA was extracted from the pure culture performing spin column purification and whole genome sequencing was conducted using Illumina's MiSeq 300PE-paired-end shotgun sequencing. The reads were *de novo* assembled, followed by genome annotation using the NCBI Prokaryotic Genome Annotation Pipeline 2.0. In addition, five *Bifidobacterium longum* subsp. *infantis* strains, available in the NCBI repositories, were selected for functional comparison. Genomic, structural, and biochemical information about the CAZymes was determined by applying the dbCAN2 and CAZy databases [3, 4]. Only genes identified from both tools were kept for downstream analysis. The obtained results were cross-examined with the UniProt protein BLAST with an E-value cut-off set to $<1e^{-10}$ and a filter with a query percentage identity value greater than 45% to receive top hits for the CAZymes. Gene clusters dedicated to the metabolism of a specific carbohydrate were identified using the dbCAN-PUL data repository [5, 6].

Functional comparison of carbohydrate utilization genes revealed notable differences between bovine and human *Bifidobacterium* strains. Genes for two glycosyl hydrolase families, GH16 and GH146, were identified, which are exclusive to the bovine strains. The endo-β-glucanase GH16 is responsible for mucin-degradation, which could influence host adaption, and the β-L-arabinofuranosidase GH146 is involved in the degradation of potential prebiotic oligosaccharides. Furthermore, CAZyme gene clusters involved in the utilization of fucosyllactose and galactooligosaccharides, which could be beneficial for an intact cattle gut function, were identified. Further studies are needed for the comprehensive understanding of prebiotic oligosaccharides and the benefit of bovine *Bifidobacterium* strains as effective probiotics, to promote the gut health of calves.

REFERENCES

1. https://doi.org/10.3389/fvets.2015.00036

2. https://doi.org/10.1007/s12263-010-0206-6
3. https://doi.org/10.1093/nar/gky418
4. https://doi.org/10.1093/nar/gkab1045
5. https://doi.org/10.1093/nar/gkac1052
6. https://doi.org/10.1093/nar/gkaa742

Identifying ac^{4}C RNA bases in nanopore direct RNA sequencing data in human brain tissues

AUTHORS
Li Wu, Luke Saville, Alexey Sizykh, Liam Mitchell, Matthew Stuart-Edwards, Athanasios Zovoilis
[1] Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, Canada
[2] Southern Alberta Genome Sciences Centre, Lethbridge, Alberta, Canada

Presenting author: li.wu@uleth.ca

KEYWORDS
RNA modification, RNA acetylation, N4-acetylcytiding, nanopore direct RNA sequencing, epitranscriptome

ABSTRACT
N4-acetylcytidine (ac^{4}C) refers to the addition of an acetyl group to cytidine and has been discovered in various RNA molecules, such as tRNA, rRNA and mRNA. Like other types of epigenetic modifications, ac4C is reversible and does not alter the original DNA sequence in the cell. Compared with the more investigated RNA methylation, studies focusing on ac4C are relatively new and still emerging. It has been shown that conserved ac4C sites exist in several tRNAs and rRNAs, and ac^{4}C are believed to have functional roles in maintaining the RNA structure stability, interacting with other molecules, and regulating gene expression. The nanopore direct RNA sequencing (DRS) technique [1] can read RNA sequences in the native form without the involvement of cDNA synthesis, making it useful in detecting RNA modifications from the voltage signal. Many analysis tools have been developed to identify such modifications. Here, we applied nanopore DRS combined with the xPore program [2] to compare chemically treated and untreated human brain tissue samples. Two types of treatments, which target modifications for acetylated and methylated nucleotides respectively, were performed. Testing a panel of neural tissues, we observed that all treated samples displayed an overall higher modification rate. Many tissues also displayed different profiles, which can later be used for constructing tissue-specific modification atlas. Our findings demonstrate the proof of principle of identifying ac4C RNA bases in nanopore DRS data, and the results can further contribute to refining the direct base calling algorithm used by nanopore sequencing.

REFERENCES
1. https://www.nature.com/articles/nmeth.4577
2. https://www.nature.com/articles/s41587-021-00949-w

Functional genomic characterization of naturalized *Escherichia coli* strains adapted to food- and water-associated industrial environments

AUTHORS: Daniel Yu [1, 2], Shuai Zhi [3, 4], and Norman F. Neumann [1, 2]

[1] School of Public Health, University of Alberta, Edmonton, Alberta, Canada
[2] Antimicrobial Resistance – One Health Consortium, Calgary, Alberta, Canada
[3] School of Medicine, Ningbo University, Ningbo, China
[4] The Affiliated Hospital of the Medical School, Ningbo University, Ningbo, China

Presenting author: dyu4@ualberta.ca

KEYWORDS

naturalized *Escherichia coli*, bacterial genome-wide association studies, environmental microbiology, meat processing plants, wastewater treatment plants

ABSTRACT

BACKGROUND: Although *Escherichia coli* is classically known as a commensal gut microbe, select strains have evolved to survive in non-host environments. These 'naturalized' *E. coli* have been identified primarily within natural contexts such as soil, sand and water (1); however, recent studies have described strains that appear to thrive in water-associated industrial environments (2,3). To understand the evolutionary emergence of these naturalized strains, we describe the functional adaptations of *E. coli* strains isolated from meat processing and wastewater treatment plants.

METHODS: The genomes of 37 previously-described naturalized wastewater and meat plant *E. coli* isolates were downloaded from NCBI alongside representative commensal *E. coli*, intestinal and extraintestinal pathogenic *E. coli*, environmental *E. coli*, and cryptic *Escherichia* strains. Roary (4) was used to estimate a pan-genome for all strains analyzed, which was then statistically scored with Scoary (5) to identify ecologically-relevant genes that statistically correlated ($p < 1E-5$, with Benjamini-Hochberg correction) with the naturalized strains. All identified, annotated genes were then functionally characterized using the UniProt (6) and EcoCyc (7) databases.

RESULTS: Pan-genome analyses estimated a pan-genome of 31335 genes, of which 2501 were found to be correlated with the naturalized wastewater strains. Of these, 2082 were found to be statistically over-represented in the naturalized strains, with functions relevant for survival in their industrial niches, including biofilm formation, microbial defense mechanisms (i.e., toxin-antitoxin systems, restriction-modification systems, and anti-phage defense systems), and stress resistance against DNA-damaging stimuli, oxidative stress, heat shock, heavy metal stress, and generalized stress conditions. Interestingly, and seemingly reflecting their naturalized status, the meat plant and wastewater strains were also found to simultaneously lack various key colonization factors, virulence genes, and metabolic genes associated with the intestinal environment.

CONCLUSIONS: Naturalized strains of *E. coli* appear to have emerged within meat processing and wastewater treatment plants, illustrating the strong selection pressures that these industrial environments may exert over the course of microbial evolution. As demonstrated by these naturalized strains, microbes facing the extreme stressors experienced within food- and water-associated industrial environments may be driven to acquire various disinfection-related stress resistance, inter-microbial competition and anti-phage defense mechanisms – adaptations that, while vital for survival in industrial contexts, seem to have come at the cost of fitness within the original host environment.

REFERENCES

1. https://doi.org/10.1111/jam.13468
2. https://doi.org/10.1016/j.watres.2019.03.024
3. https://doi.org/10.1128/AEM.01574-21
4. https://doi.org/10.1093/bioinformatics/btv421
5. https://doi.org/10.1186/s13059-016-1108-8
6. https://doi.org/10.1093/nar/gkw1099
7. https://doi.org/10.1128/ecosalplus.ESP-0006-2018

Genomic characterization of carbapenem resistant bacteria from beef cattle feedlots

AUTHORS

Sani-e-Zehra Zaidi [1,2], Rahat Zaheer[1], Krysty Thomas[1], Sujeema Abeysekara[1], Travis Haight[2], Luke Saville[2], Matthew Stuart-Edwards[2], Athanasios Zovoilis[2] & Tim A. McAllister[1]

AFFILIATIONS

1. Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, 5403 1st Avenue South, Lethbridge, AB, T1J 4P4, Canada.

2. Department of Chemistry and Biochemistry, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta T1K 3M4

KEYWORDS

Cattle production, carbapenem resistance, comparative, whole genome sequencing, genomic analysis

ABSTRACT

Carbapenems are considered a last resort for the treatment of multi-drug resistant bacterial infections in humans. In this study, we investigated the occurrence of carbapenem resistant bacteria in feedlots in Alberta, Canada. The presumptive carbapenem resistant isolates (n=116) recovered after ertapenem enrichment were subjected to antimicrobial susceptibility testing against 12 different antibiotics including four carbapenems. Of these 72% of the isolates (n=84) showed resistance to ertapenem while 27% of the isolates (n=31) were resistant to at least one other carbapenem with all except one isolate being resistant to at least two other drug classes. Of these 31 isolates, 90% were carbapenemase positive, while a subset of 36 ertapenem-only resistant isolates were carbapenemase negative. The positive isolates belonged to three genera; *Pseudomonas*, *Acinetobacter*, and *Stenotrophomonas* with the majority being Pseudomonas aeruginosa (n=20) as identified by 16S rRNA gene sequencing. Whole genome sequencing identified intrinsic carbapenem resistance genes including *blaOXA-50* and its variants (*P. aeruginosa*), *blaOXA-265 (A. haemolyticus*), *blaOXA-648 (A. lwoffii*), *blaOXA-278* (*A. junii*), and *blaL1* and *blaL2* (*S. maltophilia*). The acquired carbapenem resistance gene (*blaPST-2*) was identified in *P. saudiphocaensis* and *P. stutzeri*. In a comparative genomic analysis, clinical *P. aeruginosa*, clustered separately from those recovered from bovine feces. In conclusion, despite the use of selective enrichment methods, finding carbapenem resistant bacteria within feedlot environment is a rarity.

The Effect of Combined Treatment of Psilocybin and Eugenol on Lipopolysaccharide-Induced Brain Inflammation in Mice

AUTHORS

Presenting author: zanikovt@uleth.ca

Timur Zanikov [1,†], Marta Gerasymchuk [1,†], Esmaeel Ghasemi Gojani [1,†], Gregory Ian Robinson [1,†], Shima Asghari [1], Alyssa Groves [1], Lucie Haselhorst [1,2], Sanjana Nandakumar [1,3], Cora Stahl [1,4], Mackenzie Cameron [1], Dongping Li [1], Rocio Rodriguez-Juarez 1,5, Alexandra Snelling [5], Darryl Hudson [5], Anna Fiselier [1,6], Olga Kovalchuk [1,*] and Igor Kovalchuk [1,*]

1. Department of Biological Sciences, University of Lethbridge, Lethbridge, AB T1K 3M4, Canada
2. Institute for Medical Nutrition Science, Universität zu Lübeck, 23562 Lübeck, Germany
3. School of Biosciences and Technology, Vellore Institute of Technology, Vellore 632014, India
4. Department of Medicine, Medical Sciences, and Nutrition, University of Aberdeen, King's College, Aberdeen AB24 3FX, UK
5. GoodCap Pharmaceuticals, 520 3rd Avenue SW, Suite 1900, Calgary, AB T2P 0R3, Canada
6. Department of Family Medicine, Cumming School of Medicine, University of Calgary, Calgary, AB T2N 4N1, Canada

* Authors to whom correspondence should be addressed.

† These authors contributed equally to this work.

KEYWORDS

Psilocybin, Eugenol, LPS, Inflammation, Brain

ABSTRACT

The term "neuroinflammation" refers to an inflammatory response within the brain or spinal cord. The prevention and treatment of inflammation have been one of the main focuses of many researchers. The effect of psychedelic mushrooms on inflammation have been the object of interest of recent studies. We hypothesized psilocybin and/or eugenol would ameliorate lipopolysaccharide (LPS)-induced inflammation. Inflammation was induced by intraperitoneal injection of LPS. Following LPS treatment, mice were treated with psilocybin, a main component of psychedelic mushrooms, and eugenol, derived from cloves, or their combination. Mice were euthanized and brains were collected and frozen. Western Immunoblotting and reverse transcription-quantitative polymerase chain reaction (RT-qPCR) were utilized to analyze expression level of pro-inflammatory cytokines and receptor proteins in mice's brain. Our data suggests that eugenol and some combinations with psilocybin have beneficial effects on treating neuroinflammation. This study is the first, which demonstrates the anti-inflammatory effects of combined treatments of eugenol and psilocybin on neuroinflammation.

User-friendly genomic neighbourhood visualization tool

AUTHORS
Nolen Timmerman[1], Chantel Trost[2], Alan R. Davidson[2], Lingling Jin[1*]

[1] University of Saskatchewan, Saskatoon, Canada
[2] University of Toronto, Toronto, Canada

Presenting author: njt694@usask.ca

KEYWORDS
Anti-CRISPR, CRISPR, bacteriophage, genomic neighbourhood, visualization

ABSTRACT
Bacteria and bacteriophages (phages) are involved in an evolutionary arms race, as phages infect and destroy bacteria and bacteria must overcome phage attack in order to survive. To defend against incoming phages, bacteria have evolved numerous defense systems, including the adaptive Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated proteins (CRISPR-Cas) systems. In response, phages have evolved anti-CRISPR (Acr) proteins, which are able to inhibit CRISPR-Cas systems. Acrs are typically very small proteins that are highly variable in both sequence and structure; thus, they are difficult to identify bioinformatically using traditional, sequence-based methods. However, *acr* genes tend to cluster near each other, and are frequently positioned upstream of anti-CRISPR associated (*aca*) genes, which encode for highly conserved proteins that regulate *acr* gene expression. This observation has led to the guilt-by-association method of Acr identification, in which the genomic neighbourhoods of known *acr* and *aca* genes are used to search for new ones, which can then be used to discover more. Key caveats of this approach are the requirements for small databases to avoid overwhelming the end-user, the inability to intuitively visualize the matches, and the difficulty in interpreting unannotated proteins in the searched genomes. To overcome these limitations, we developed a bioinformatics pipeline to rapidly find and visualize the genomic neighbourhoods of *acr* genes. The pipeline searches a large database of bacterial or viral genomes using PSI-BLAST for matches to the Acr query, clusters genomic neighborhoods containing query matches that are similar, and then generates visually appealing diagrams showing the genomic neighbourhoods of the matches. It also outputs taxonomic summaries of the PSI-BLAST results. Although this pipeline was designed for analyzing the genomic neighbourhoods of anti-CRISPR genes, it functions with any protein sequence as input, and thus can help visualize the genomic neighbourhood of any protein under study.

ModSpy: A Machine Learning model detects Genetic Modifiers from Whole Genome Sequencing data of Model Organisms

AUTHORS:
K M Tahsin Hassan Rahit[1, 2], Maja Tarailo-Graovac[1, 2]

[1] Departments of Biochemistry, Molecular Biology and Medical Genetics, Cumming School of Medicine, University of Calgary, AB T2N 4N1, Canada
[2] Alberta Children's Hospital Research Institute, University of Calgary, Calgary, AB T2N 4N1, Canada

Presenting author: kmtahsinhassan.rahit@ucalgary.ca

KEYWORDS
Genetic Modifier, Machine Learning, Whole Genome Sequencing, Suppressor, *C. elegans*, Rare Disease

ABSTRACT

Identifying genetic modifiers, which are genetic variants that affect the phenotypic outcome of another genetic variant, is essential for understanding the penetrance and expressivity of rare genetic disorders [1]. Whole Genome Sequencing (WGS) data analysis provides a powerful approach for exploring individual genomes at the variant level, but computational methods are needed to facilitate identification of genetic modifiers from WGS data.

In this study, we develop a novel computational model to expedite modifier identification from WGS data. We focus on suppressors as a specific type of modifiers that alleviate the consequence caused by a primary deleterious variant. Given the infrequency of rare disease cases, we utilize the nematode *Caenorhabditis elegans* as a model organism. We conduct high-throughput forward mutagenesis screening with *C. elegans*, followed by WGS analysis to identify potential suppressors. During WGS analysis, we utilize a combination of bioinformatics tools and manual assessment. The automated analyses typically yields a candidate variant list (CVL) of about 100 variants for each genome, which are then manually evaluated for specific features such as gene functionalities, interactions, and phenotypic similarity by human experts to select the best candidate variant. Lastly, multiple experimental approaches are employed to confirm the modifying role of the candidate variant.

In our effort to reduce the bottlenecks associated with manual analyses, we developed a machine learning (ML) model called ModSpy to prioritize suppressor variants based on underlying biological mechanisms. We trained and tested the model using *zyg-1* suppressor screening data. ZYG-1 is an ortholog of PLK4 in human, which has been associated with microcephaly. We validated the model using WGS data beyond *zyg-1*, including independently published *nekl-2* and *nekl-3* suppressor screening data. In all cases, ModSpy model has reduced the candidate list by 90%. In case we consider top-five ranked gene; it shows an accuracy of 85%.

Our findings demonstrate the potential for broad applicability of ModSpy in identifying suppressors from WGS data. Our approach could be extended to other types of genetic modifiers (i.e., modifier networks). Additionally, our approach highlights the power of ML in combining multiple sources of evidence and reducing the expert time spent on manual analysis of CVLs aiding in high throughput design of our screens. This can facilitate more effective utilization of model organisms in search of modifiers relevant to human

disease, which is crucial for understanding the mechanisms of disease, variable presentation of and may reveal therapeutic targets.

REFERENCES

1. https://doi.org/10.3390/genes11030239

Proksee: a web server for the characterization and visualization of bacterial genomes.

AUTHORS

List of authors (Underline the name of presenting author and provide an e-mail address), Institute Name, Town, Country

Jason R. Grant[1], Eric Enns[2], Eric Marinier[2], Arnab Saha-Mandal[3], Emily K. Herman[1], Chih-Yu Chen[2], Morag Graham[2], Gary Van Domselaar[2], Paul Stothard[1]

[1] University of Alberta, Edmonton, Canada

[2] National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Canada

[3] University of Manitoba, Winnipeg, Canada

Presenting author: jason.grant@ualberta.ca

KEYWORDS

bacteria, genome maps, visualization, annotation, assembly

ABSTRACT

Proksee (https://proksee.ca) is a robust and user-friendly platform that allows for the assembly, analysis, and visualization of bacterial genomes. It accommodates a wide range of input types including raw sequencing reads (FASTQ files), preassembled contigs (FASTA, GenBank or EMBL files), GenBank accessions, or previously generated Proksee maps in JSON format. For raw sequence data, Proksee performs assembly and offers a comparison of assembly metrics using a custom reference database of assemblies from the same species. All projects present a graphical map of the bacterial genome using the custom-built CGView.js browser, which is optimized for circular genomes, and enables quick navigation and zooming down to the sequence level. The map is accompanied by a collection of panels for customizing it, examining features, initiating new analysis jobs, and downloading results. Each analysis job is equipped with a real-time log, a built-in file viewer, and a report that describes important files, provides links to helpful resources, and can seamlessly integrate results into the map. Proksee offers a growing number of tools for various analyses, such as genome annotation (e.g. Prokka, CARD RGI, CRISPR/Cas Finder), identification of mobile genetic elements (e.g. VirSorter, Phigaro, mobileOD-db), and sequence comparison (e.g. BLAST, FastANI). These computationally intensive analyses run as jobs on a multi-server cloud-based system that forms the backend of Proksee. An integrated admin system facilitates management of the system, providing real-time updates on user numbers, server load, data submission, running jobs, and job errors. Proksee is designed to be highly extensible and over the next year several significant enhancements are planned.

Copper Sulphate Foliar Applications and Soil Amendments Alters Pathogens Found on Cannabis sativa

AUTHORS
Gregory Ian Robinson[1,2], Andreas Eriksson[1], Hamed Farahmand[1], Marta Gerasymchuk[1], Igor Kovalchuk[1]
[1] University of Lethbridge, Lethbridge, Canada
[2] University of Illinois Urbana-Champaign, Urbana, United States of America

Presenting author: g.robinson@uleth.ca

KEYWORDS
Pest Management, Cannabis, Penicillium, Copper Sulphate, Bioinformatics

ABSTRACT
Fungal infections of *Cannabis sativa* are common and can severely affect yields, however, few studies have examined the use of fungicides for cannabis production. Copper (II) sulphate is an age-old fungicide that has been used to impede fungal diseases in multiple crops while boosting plant defense systems through increased endogenous ethylene biosynthesis. This study aims to test the efficacy of copper sulphate against cannabis pathogens. Copper sulphate was added as a soil amendment from rooting or applied as a foliar application to 10-week-old *C. sativa* plants biweekly for four weeks. Leaves were collected, frozen, and genomic DNA was isolated and sequenced. Overall pathogen count was decreased in plants treated with copper sulphate as a soil amendment (1 – 10 mg/L), however, plant growth was unaffected. Divergent results of the abundance of bacteria, fungi, and viruses were seen between soil types and chemotypes of *C. sativa*. Over 300 pathogens were enriched, while under 200 pathogens were depleted with copper sulphate treatment. In contrast, foliar applications (100 mg/L) inhibited the presence of *Penicillium olsonii,* which causes penicillium bud rot, an important post-harvest disease in *C. sativa,* and decreased conidiogenesis on all cultivars tested. To verify the fungicidal activity of copper sulphate, *P. olsonii* was grown *in vitro* on potato dextrose agar supplemented with various concentrations of copper sulphate (0 - 300 mg/L). The half maximal effective concentration (EC_{50}) and minimum inhibitory concentration (MIC) was calculated to be 73.1 mg/L and 474.9 mg/L, respectively. This is the first report showing soil amendments and foliar applications of copper sulphate on cannabis can alter pathogen abundance and prevent fungal disease.

Transcriptomic patterns of PRRSV infection: potential biomarkers and further research

AUTHORS
Aleksei Sizykh, Athanasios Zovoilis
[1] University of Lethbridge, Lethbridge, Canada

Presenting author: Alexey.sizykh@uleth.ca

KEYWORDS
SINEs, retrotransposons, repetitive éléments, swine, bioinformatics

ABSTRACT
Porcine Reproductive Respiratory Syndrome (PRRS) caused by the corresponding Porcine Reproductive Respiratory Syndrome Virus (PRRSV) is one of serious diseases damaging the pig production: it significantly reduces the litter viability and limits swine growth. Research of mechanisms, patterns and markers of this infection is necessitated by possible positive economical effects on the domestic livestock industry. Short interspersed nuclear elements (SINEs). A family of retrotransposons known to be silent in healthy cells but expressed under stress, might play a role in this disease. The role of Porcine repetitive element 1 (PRE1), the most abundant SINE in pigs, in PRRSV is investigated in this study.

Epitranscriptomic Assessment of N4-Acetylcytidine and N6-Methyladenosine in Mouse Neural Tissues.

AUTHORS
Luke Saville, Li Wu, Alexey Sizykh, Liam Mitchell, Matthew Stuart-Edwards, Athanasios Zovoilis
[1] University of Lethbridge, Lethbridge, Canada

Presenting author: luke.saville@uleth.ca

KEYWORDS
Neural, N4-acetylcytidine, N6-Methyladenosine, RNA modification, Nanopore sequencing

ABSTRACT
The epitranscriptome is a fine-tunable and dynamic system of RNA modifications, appearing in diverse forms across genera. Because of limitations in detection methods, the functional annotation of RNA modifications has been limited. While assessment of quantities of RNA modifications and even the reads in which those RNA modifications may preside has long been a capability of RNA researchers, only recently has site specific detection been possible. Furthermore, detection of modifications has largely been limited to assessing a single RNA modification at a time. Oxford Nanopore Technologies (ONT)'s development of direct RNA sequencing poses immense potential for site-specific detection of any present RNA modification whereby, the RNA molecule is sequenced in its native form and disturbs an ion gradient through a nanoscale protein pore, detected as current. Discrete aberrations in current are thought to be potential RNA modifications and by producing comparative datasets, the site-specific signal aberrations can be attributed. Herein, we present the assessment of N4-acetylcytidine (ac4C) and N6-methyladenosine (m6A) by producing chemically de-modified comparative sets in differing mouse neural tissues and comparisons with ncRNAs *in vitro* and in an Alzheimer's disease model context.