Laboratory of Sudhir Kumar

January 2021

kumarlab.net

Lab Members

Postdoctoral & Doctoral
Caryn Babaian
Jose Barbo-Montoya
Marcos Caraballo-Ortiz
Antonia Chroni
Jack Craig
Sayaka Miura
Ravi Patel
Lisa Schmelkin
Sudip Sharma
Qiqing Tao

Grad/Undergrad & Intern
Vivian Aly
Jiyeong Choi
Lauren Hamilton
Abu Hasnat Hasib
Jared Huzar
Julia Davis
Nisarg Patel
Tina To
Sarah Vahdatmoar
Tracy Vu

Technologists & Staff
Jared Knoblauch
Michael Li
Heather Rowe
Maxwell Sanderford
Nick Harris
Glen Stecher
Michael Suleski

Books

Masatoshi Nei &
Sudhir Kumar

S. Blair Hedges &
Sudhir Kumar

Education Initiatives

We are developing undergraduate degree and certificate programs, participating in masters’ and doctoral degree programs, offering advanced courses, and providing one-on-one research mentoring to undergraduate, graduate, and postgraduate students.

B.S. in Genomic Medicine
B.S. in Data Science (Genomics and Bioinformatics)
Undergrad Certificate in Genomic Medicine
Genomic Evolutionary Medicine
PathFinder: Bayesian inference of clone migration histories in cancer *Bioinformatics*

Power and pitfalls of computational methods for inferring clone phylogenies and mutation orders from bulk sequencing data *Scientific Reports* | 10:3498

Computational enhancement of single-cell sequences for inferring tumor evolution *Bioinformatics* | 34:i917

Predicting clone genotypes from tumor bulk sequencing of multiple samples *Bioinformatics* | 34:4017

Somatic evolutionary timings of driver mutations *BMC Cancer* | 18:85

Molecular Phylogenetics

**Relative efficiencies of simple and complex substitution models in estimating divergence times in phylogenomics** *MBE* | 37:1819

**A new method for inferring timetrees from temporally sampled molecular sequences** *PLoS Computational Biology* | 16:e1007046

**Reliable confidence intervals for RelTime estimates of evolutionary divergence times** *MBE* | 37:280

**A machine learning method for detecting autocorrelation of evolutionary rates in large phylogenies** *MBE* | 36:811

**Theoretical foundation of the RelTime method for estimating divergence times** *MBE* | 35:1770

**Phylogenetic placement of metagenomic reads using the minimum evolution principle** *BMC Genomics* | 16:S13

**Prospects for building large timetrees using molecular data with incomplete gene coverage among species** *MBE* | 31:2542

**Estimating divergence times in large molecular phylogenies** *PNAS* | 109:19333

**Prospects for inferring very large phylogenies by using the neighbor-joining method** *PNAS* | 101:11030

**Disparity index: a simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences** *Genetics* | 158:1321

**Incomplete taxon sampling is not a problem for phylogenetic inference** *PNAS* | 98:10751

**Stepwise algorithm for finding minimum evolution trees** *MBE* | 13:584

**Four-cluster analysis: a simple method to test phylogenetic hypotheses** *MBE* | 12:163

**Tree of life reveals clock-like speciation and diversification** *MBE* | 32:835

**Placing confidence limits on the molecular age of the human-chimpanzee divergence** *PNAS* | 102:18842

**Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks** *MBE* | 21:36

**A molecular timescale for vertebrate evolution** *Nature* | 392:917

**Continental breakup and the ordinal diversification of birds and mammals** *Nature* | 381:226

**Adventures in Evolution: The Narrative of Tardigrada, Trundlers in Time** *ABT* | 81:543

**Human Molecular Genetics**

**Understanding human disease mutations through the use of interspecific genetic variation** *Human Molecular Genetics* | 10:2319

**Continental breakup and the ordinal diversification of birds and mammals** *Nature* | 381:226

**Understanding human disease mutations through the use of interspecific genetic variation** *Human Molecular Genetics* | 10:2319

**PathFinder: Bayesian inference of clone migration histories in cancer** *Bioinformatics*

**Power and pitfalls of computational methods for inferring clone phylogenies and mutation orders from bulk sequencing data** *Scientific Reports* | 10:3498

**Computational enhancement of single-cell sequences for inferring tumor evolution** *Bioinformatics* | 34:i917

**Predicting clone genotypes from tumor bulk sequencing of multiple samples** *Bioinformatics* | 34:4017

**Somatic evolutionary timings of driver mutations** *BMC Cancer* | 18:85

**Molecular Memories of a Cambrian Fossil** *ABT* | 82:586

**Adventures in Evolution: The Narrative of Tardigrada, Trundlers in Time** *ABT* | 81:543

**An evolutionary portrait of the progenitor SARS-CoV-2 and its dominant offshoots in COVID-19 pandemic** *bioRxiv*

**Biological relevance of computationally predicted pathogenicity of noncoding variants** *Nature Communications* | 10:330

**Evolutionary Diagnosis of non-synonymous variants involved in differential drug response** *BMC Medical Genomics* | 8:S6

**Genome-wide profiling of human cap-independent translation-enhancing elements** *Nature Methods* | 10:747

**Performance of computational tools in evaluating the functional impact of laboratory-induced amino acid mutations** *Bioinformatics* | 28:2093

**Evolutionary diagnosis method for variants in personal exomes** *Nature Methods* | 9:855

**Positional conservation and amino acids shape the correct diagnosis and population frequencies of benign and damaging personal amino acid mutations** *Genome Research* | 19:1562

**Understanding human disease mutations through the use of interspecific genetic variation** *Human Molecular Genetics* | 10:2319
Reviews & Perspectives

- **HOT PAPERS**
  - Among the top 0.1% of all the papers published in two years following their release
    - Bioinformatics 17:1244
    - Nature 392: 917
    - MBE 35:1547 and 34:1812
    - Briefings in Bioinformatics 9:299 and 5:150

- **HIGHLY CITED**
  - Citation count in the 1% of all articles in a 10-year period
    - PNAS, USA 101:11030
    - MBE 35:1547; 34:1812; 33:1870; 32:835; 30:2735–2729; 28:2731; and 24:1596

- **TOP CITED**
  - In Nature’s top-100 articles of all time in the published literature
    - MBE 24:1596

---

**YEAR**

**SELECTION**

- '19 On estimating evolutionary probabilities of population variants *BMC Evolutionary Biology* 19:133
- '16 A molecular evolutionary reference for the human variome *MBE* 33:245
- '04 Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome *Genetics* 168:373
- '95 A new method of inference of ancestral nucleotide and amino-acid-sequences *Genetics* 141:1641

**STRUCTURE**

- '15 Conformational dynamics of nonsynonymous variants at protein interfaces reveals disease association *Proteins* 83:428
- '14 Signatures of natural selection on mutations of residues with multiple posttranslational modifications *MBE* 31:1641
- '13 Structural dynamics flexibility informs function and evolution at a proteome scale *Evolutionary Applications* 6:423

**DEVELOPMENT**

- '16 Deep model based transfer and multi-task learning for biological image analysis *Transactions on Big Data*
- '12 Comparison of embryonic expression within multigene families using the FlyExpress discovery platform reveals more spatial than temporal divergence *Developmental Dynamics* 241:150
- '08 Automated annotation of Drosophila gene expression patterns using a controlled vocabulary *Bioinformatics* 24:1881
- '02 BEST: a novel computational approach for comparing gene expression patterns from early stages of Drosophila melanogaster development *Genetics* 162:2037

**ECOLOGY**

- '09 Signatures of nitrogen limitation in the elemental composition of the proteins involved in the metabolic apparatus *Proceedings of the Royal Society B-Biological Sciences* 276:2605
- '06 Signatures of ecological resource availability in the animal and plant proteomes *MBE* 23:1946

---

**MBE** | Molecular Biology and Evolution
**PNAS** | Proceedings of the National Academy of Sciences (USA)
**ABT** | The American Biology Teacher
MEGA

User-friendly software suite for analyzing DNA and protein sequence data from species and populations.

www.megasoftware.net

Timeline of major MEGA software releases

MEGA 1
1993

MEGA 2
2001

MEGA 5
2011

MEGA X
2018

MEGA 11
2021

Available on

windows  macOS  linux

GUI  Command Line

MEGA Publications

2020
Molecular Evolutionary Genetics Analysis (MEGA) for macOS  MBE | 37:1237

2018
MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms  MBE | 35:1547

2016
MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets  MBE | 33:1870

2014
MEGA-MD: molecular evolutionary genetics analysis software with mutational diagnosis of amino acid variation  Bioinformatics | 30:1305

2012
MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis  Bioinformatics | 28:2685

2008
MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences  Briefings in Bioinformatics | 9:299

2001
MEGA2: Molecular Evolutionary Genetics Analysis software  Bioinformatics | 17:1244

1994
MEGA - Molecular Evolutionary Genetics Analysis software for microcomputers  Computer Applications in the Biosciences | 10:189

www.timetree.org

A public knowledge-base of the evolutionary timescales of the tree of life.

TimeTree Publications

2017
TimeTree: A resource for timelines, timetrees, and divergence times  MBE | 34:1812–1819

2011
TimeTree2: species divergence times on the iPhone  Bioinformatics | 27:2023–2024

2006
TimeTree: a public knowledge-base of divergence times among organisms  Bioinformatics | 22:2971–2972

Flyexpress

www.flyexpress.net

Integrated discovery platform to study coexpressed genes using in situ hybridization images in Drosophila.

myPEG

www.mypeg.info

Integrated platform of evolutionary tools for phylomedicine.