Alexander Kozik

Current occupation:
Bioinformatics Specialist
Genome Center (Genome and Biomedical Sciences Facility)
451 Health Sciences Drive, University of California; Davis, CA 95616-8816

Contact info:
Web: *http://www.atgc.org/*
Email: *akozik@atgc.org*
Cell: (530)-220-3385

Education:
- PhD in Genetics. 1996. Wageningen Agricultural University, The Netherlands.
- M.S. in Chemistry. 1987. St Petersburg State University, Russia

Professional areas:
- Bioinformatics, large-scale data analysis, visualization, and presentation.
- Developing strategies for large-scale data handling and processing.
- Algorithm development, including testing and troubleshooting.
- Data validation and extracting meaningful information from noisy data.
- Processing and analyzing DNA sequences for next-generation sequencing platforms.
- Whole genome and transcriptome assemblies, analysis, and visualization.
- Exploring, evaluation and developing novel solutions for analyzing new types of genomic data.

Professional history:

● **2004 – current   Bioinformatics Specialist, UC Davis Genome Center.**
Genome assembly and analysis using next-generation (Illumina) sequencing technologies. Completion of lettuce transcriptome and gene space assembly. Management of 20 TB of genomics data on the departmental file server. Development and support of MadMapper, used to construct and visualize ultra-high dense genetic maps. Completion of lettuce genetic map construction with 12,000 markers. Development and support of bioinformatics tools on Google Code. Data submission to NCBI GenBank.

● **2000 – 2006   Allometra LLC.**
Design and development of PyMood, a desktop application for visualizing genomic data: http://www.allometra.com/. Implementation of BLAST launcher and parser. Design of PyMood's underlying data structures and visualization algorithms.

● **2000 – 2004   Postdoctoral fellow, Department of Vegetable Crops, UC Davis.**
Development, publishing, and support of genome analysis and visualization tools: GenomePixelizer, GenoPix 2D Plotter, and PhyloGrapher. Design and management of a database with one million DNA sequences http://compgenomics.ucdavis.edu/compositae_index.php. Development and implementation of tools to find, analyze, and visualize polymorphism in DNA within related species.

● **1996 – 2000   Postdoctoral fellow, UC Davis School of Medicine.**
Technical and programming support of lab robotic devices (Biomek-2000, Multimec-96) and DNA analyzer ABI Prism 3700. Development of protocols for high-throughput genotyping on human samples. Installation and management of the Phred/Phrap/Consed suite for DNA sequence analysis.

● **1991 – 1996   PhD student, Wageningen Agricultural University.**
Plant-microbe interaction, host-specific interaction between plants and nitrogen-fixing bacteria. Construction of fine genetic maps of plant genes involved in symbiosis with bacteria.

Programming languages:

- **Python (expert level)**
- **Tcl/Tk (expert level)**
- PHP (basic level)
- Perl (basic level)
- R (basic level)
- bash shell scripting (basic level)

Software developed:

- PyMood - 3D visualization tool for comparative genomics:
  http://allometra.com/

- CheckMatrix and MadMapper - for genetic map validation and visualization:
  http://www.atgc.org/XLinkage/MadMapper/

- GenomePixelizer - for visualizing genome organization:
  http://www.atgc.org/GenomePixelizer/

- GenoPix2D plotter - for visualization similarities between genomes:
  http://www.atgc.org/GenoPix_2D_Plotter/

- PhyloGrapher - a graph-drawing tool for biological data:
  http://www.atgc.org/PhyloGrapher/

- TkLife - a multipurpose visualization program:
  http://www.atgc.org/TkLife/

- SNP/InDel discovery and analysis - for finding genetic differences between genomes:
  http://cgpdb.ucdavis.edu/SNP_Discovery_CDS/

- Scripts and tools to process and analyze NCBI BLAST related data:
  http://code.google.com/p/atgc-tools/

- SeqExter: DNA string search and alignment – consensus building by extension:
  http://code.google.com/p/atgc-wint/

- Illumina pipeline – QC, filtering and *de novo* assembly:
  http://code.google.com/p/atgc-illumina/

Publications:

Lai Z, Kane NC, Kozik A, Hodgins KA, Dlugosch KM, Barker MS, Matvienko M, Yu Q, Turner KG, Pearl SA, Bell GD, Zou Y, Grassa C, Guggisberg A, Adams KL, Anderson JV, Horvath DP, Kesseli RV, Burke JM, Michelmore RW, Rieseberg LH.
Genomics of Compositae weeds: EST libraries, microarrays, and evidence of introgression.
Am J Bot. 2011 Nov 4. [Epub ahead of print] [PubMed: 22058181]

Cantu D, Govindarajulu M, Kozik A, Wang M, Chen X, Kojima KK, Jurka J, Michelmore RW, Dubcovsky J.
Next generation sequencing provides rapid access to the genome of Puccinia striiformis f. sp. tritici, the causal agent of wheat stripe rust.
PLoS One. 2011;6(8):e24230. Epub 2011 Aug 31 [PubMed: 21909385]

Cabrera A, Kozik A, Howad W, Arus P, Iezzoni AF, van der Knaap E.
Development and bin mapping of a Rosaceae Conserved Ortholog Set (COS) of markers.
BMC Genomics. 2009 Nov 29;10:562 [PubMed: 19943965]

Van Deynze A, Stoffel K, Lee M, Wilkins TA, Kozik A, Cantrell RG, Yu JZ, Kohel RJ, Stelly DM.
Sampling nucleotide diversity in cotton.
BMC Plant Biol. 2009 Oct 20;9:125 [PubMed: 19840401]

Wroblewski T, Caldwell KS, Piskurewicz U, Cavanaugh KA, Xu H, Kozik A, Ochoa O, McHale LK, Lahre K, Jelenska J, Castillo JA, Blumenthal D, Vinatzer BA, Greenberg JT, Michelmore RW.
Comparative large-scale analysis of interactions between several crop species and the effector repertoires from multiple pathovars of Pseudomonas and Ralstonia.
Plant Physiol. 2009 Aug;150(4):1733-49. Epub 2009 Jul 1. [PubMed: 19571308]

Barker MS, Kane NC, Matvienko M, Kozik A, Michelmore RW, Knapp SJ, Rieseberg LH.
Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years.
Mol Biol Evol. 2008 Nov;25(11):2445-2455. Epub 2008 Aug 26. [PubMed: 18728074]

Heesacker A, Kishore VK, Gao W, Tang S, Kolkman JM, Gingle A, Matvienko M, Kozik A, Michelmore RM, Lai Z, Rieseberg LH, Knapp SJ.
SSRs and INDELs mined from the sunflower EST database: abundance, polymorphisms, and cross-taxa utility.
Theor Appl Genet. 2008 Nov;117(7):1021-1029. Epub 2008 Jul 17. [PubMed: 18633591]

Radwan O, Gandhi S, Heesacker A, Whitaker B, Taylor C, Plocik A, Kesseli R, Kozik A, Michelmore RW, Knapp SJ.
Genetic diversity and genomic distribution of homologs encoding NBS-LRR disease resistance proteins in sunflower.
Mol Genet Genomics. 2008 Aug;280(2):111-125 [PubMed: 18553106]

Van Deynze AE, Stoffel K, Buell RC, Kozik A, Liu J, van der Knaap E, Francis D.
Diversity in conserved genes in tomato.
BMC Genomics. 2007 Dec 18;8(1):465 [PubMed: 18088428]

Truco MJ, Antonise R, Lavelle D, Ochoa O, Kozik A, Witsenboer H, Fort SB, Jeuken MJ, Kesseli RV, Lindhout P, Michelmore RW, Peleman J.
A high-density, integrated genetic linkage map of lettuce (Lactuca spp.).
Theor Appl Genet. 2007 Oct;115(6):735-746 [PubMed: 17828385]

Timms L, Jimenez R, Chase M, Lavelle D, McHale L, Kozik A, Zhao L, Heesacker A, Knapp S, Rieseberg L, Michelmore R, Kesseli R.
Analyses of synteny between Arabidopsis thaliana and species in the Asteraceae reveal a complex network of small syntenic segments and major chromosomal rearrangements.
Genetics. 2006 Aug;173(4):2227-2235. Epub 2006 Jun 18. [PubMed: 16783026]

West MA, van Leeuwen H, Kozik A, Kliebenstein DJ, Doerge RW, St Clair DA, Michelmore RW.
High-density haplotyping with microarray-based expression and single feature polymorphism markers in Arabidopsis.
Genome Res. 2006 Jun;16(6):787-795. Epub 2006 May 15. [PubMed: 16702412]

Cannon SB, Kozik A, Chan B, Michelmore R, Young ND.
DiagHunter and GenoPix2D: programs for genomic comparisons, large-scale homology discovery and visualization.
Genome Biol. 2003;4(10):R68. [PubMed: 14519203]

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW.
Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis.
Plant Cell. 2003 Apr;15(4):809-834. [PubMed: 12671079]

Kozik A, Kochetkova E, Michelmore R.
GenomePixelizer-a visualization program for comparative genomics within and between species.
Bioinformatics 2002 18: 335-336. [PubMed: 11847088]