



Tips and Tricks for Genomic Analysis in SVS

August 6, 2014

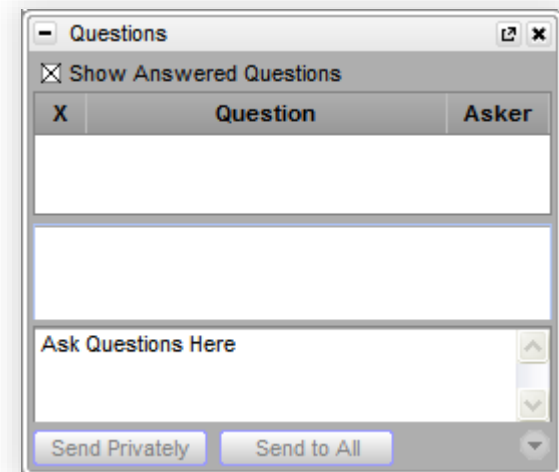
Ashley Hintz

Field Application Scientist



Questions during the presentation

Use the Questions pane in your GoToWebinar window





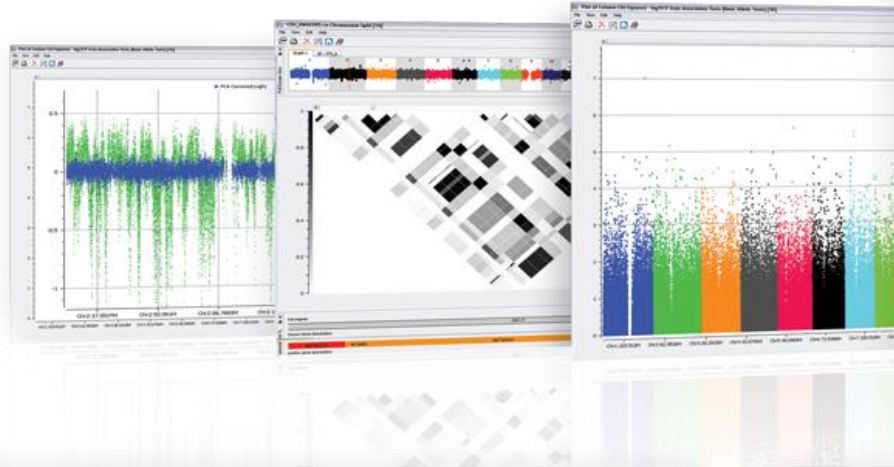
1 Tips for DNaseq

2 Data Management Tricks

3 Build Custom Genome

4 Questions!

SNP & Variation Suite (SVS)



Core Features

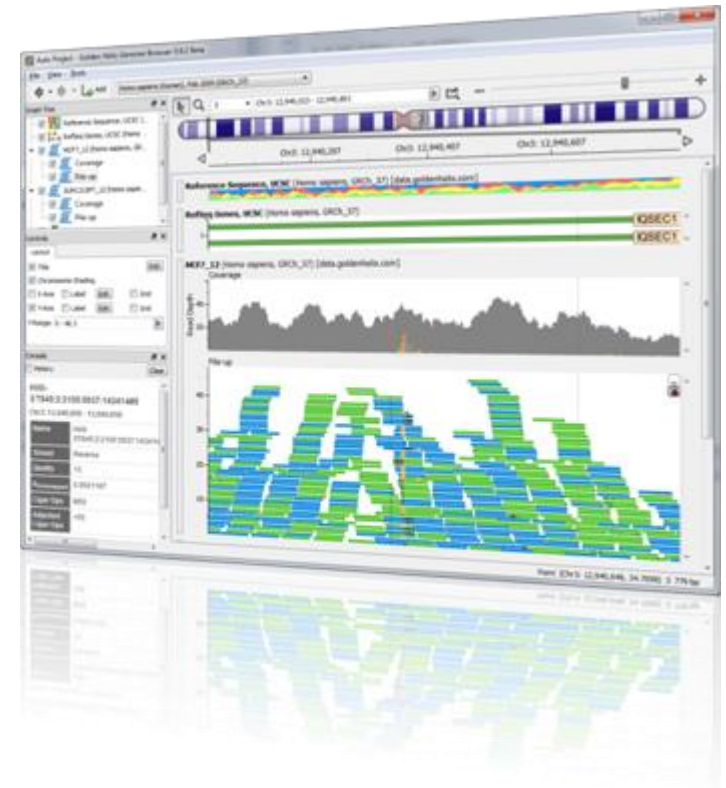
- Powerful Data Management
- Rich Visualizations
- Robust Statistics
- Flexible
- Easy-to-use

Applications

- Genotype Analysis
- DNA sequence analysis
- CNV Analysis
- RNA-seq differential expression
- Family Based Association



- Free sequencing visualization tool
- Launched in 2011
- Makes the process of exploring DNA-seq and RNA-seq pile-up and coverage data intuitive and powerful
- Stream public annotations via the cloud
- Use it to validate variant calls, trio exploration, de Novo discovery, and more





1 Tips for DNaseq

2 Data Management Tricks

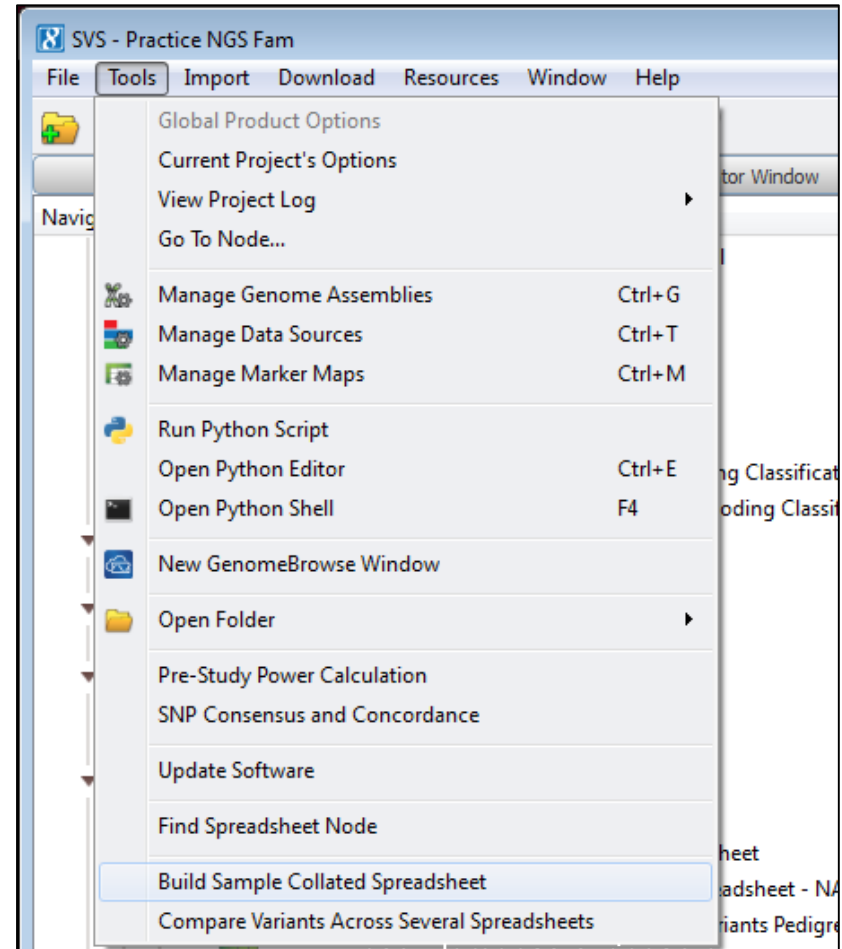
3 Build Custom Genome

4 Questions!

Tips for DNASEq



- Ability to create a collated and filtered spreadsheet from multiple inputs
 - Outputs data from where all spreadsheets overlap
- Sample use case – de Novo Variants
- Script: *Join or Merge Several Spreadsheets*





1 Tips for DNaseq

2 Data Management Tricks

3 Build Custom Genome

4 Questions!

Data Management Tricks



- Activate by Gene List
- Activate by category in data columns
- Markers in LD block subset in Genome Browse
- Script: *Activate/Inactivate based on Genomic Position*
- Script: *Inactivate Duplicate Column Headers*



The screenshot shows a spreadsheet window titled "Brain Expression Data - Sheet 1 [17]". The spreadsheet has columns for "column_num", "FAM138A", "FAM138A", "FAM138A", "FAM138F", "FAM138F", "FAM138F", "OR4F5", and "RP4-669L17.10". The rows are numbered 1 through 25. The data in the rows is as follows:

column_num	FAM138A	FAM138A	FAM138A	FAM138F	FAM138F	FAM138F	OR4F5	RP4-669L17.10
41	0	0	0	0	0	0	0	0
53	0	0	0	0	0	0	0	0
63	0	0	0	0	0	0	0	0
78	0	0	0	0	0	0	0	0
105	0	0	0	0	0	0	0	0
113	0	0	0	0	0	0	0	0
122	0	0	0	0	0	0	0	0
142	0	0	0	0	0	0	0	0
157	0	0	0	0	0	0	0	0
165	0	0	0	0	0	0	0	0
175	0	0	0	0	0	0	0	0
205	0	0	0	0	0	0	0	0
216	0	0	0	0	0	0	0	0
235	0	0	0	0	0	0	0	0
248	0	0	0	0	0	0	0	0
267	0	0	0	0	0	0	0	0
284	0	0	0	0	0	0	0	0
318	0	0	0	0	0	0	0	0
325	0	0	0	0	0	0	0	0
342	0	0	0	0	0	0	0	0
363	0	0	0	0	0	0	0	0
378	0	0	0	0	0	0	0	0
385	0	0	0	0	0	0	0	0
410	0	0	0	0	0	0	0	0
420	0	0	0	0	0	0	0	0



GOLDEN HELIX SNP & VARIATION SUITE

[Demonstration]



1 Tips for DNaseq

2 Data Management Tricks

3 Build Custom Genome

4 Questions!

Build Custom genome



- *Anopheles gambiae* (Mosquito)
- *Arabidopsis thaliana* (Thale Cress)
- *Bos taurus* (Cow)
- *Brassica rapa* (Napa Cabbage)
- *Caenorhabditis elegans* (Round Worm)
- *Canis familiaris* (Dog)
- *Capra hircus* (Goat)
- *Capsicum annuum* (Hot Pepper)
- *Carica papaya* (Papaya)
- *Citrullus lanatus* (Watermelon)
- *Cricetulus griseus* (Chinese Hamster Ovary- CHO)
- *Danio rerio* (Zebrafish)
- *Drosophila melanogaster* (Fruit Fly)
- *Equus caballus* (Horse)
- *Eucalyptus grandis* (Eucalyptus)
- *Felis catus* (Cat)
- *Gallus gallus* (Chicken)
- *Glycine max* (Soybean)
- *Gossypium raimondii* (Cotton D)
- *Heterocephalus glaber* (Naked Mole-Rat)
- *Leishmania infantum* JPCM5 (Leishmania parasite)
- *Macaca mulatta* (Rhesus Monkey)
- *Mus musculus* (Mouse)
- *Onchorhynchus mykiss* (Rainbow Trout)
- *Oryza sativa* (Japanese Rice)
- *Ovis aries* (Sheep)
- *Prunus persica* (Peach)
- *Rattus norvegicus* (Norway Rat)
- *Saccharomyces pombe* (Baker's Yeast)
- *Sorghum bicolor* (Sorghum)
- *Sus scrofa* (Pig)
- *Zea mays* (Maize)

Build Custom Genome



- *Anolis carolinensis* – Carolina Anole Lizard
- Downloaded from [Ensembl](#)
- Reference Genome
 - Separate FASTA file for each chromosome
- Gene Track
 - Single GTF file
- Originally published in Nature by Alföldi *et al.* (2011)
- 6 macro-chromosomes
 - Draft assembly 1.78GB
 - 17,472 protein-coding genes
 - 2,924 RNA genes

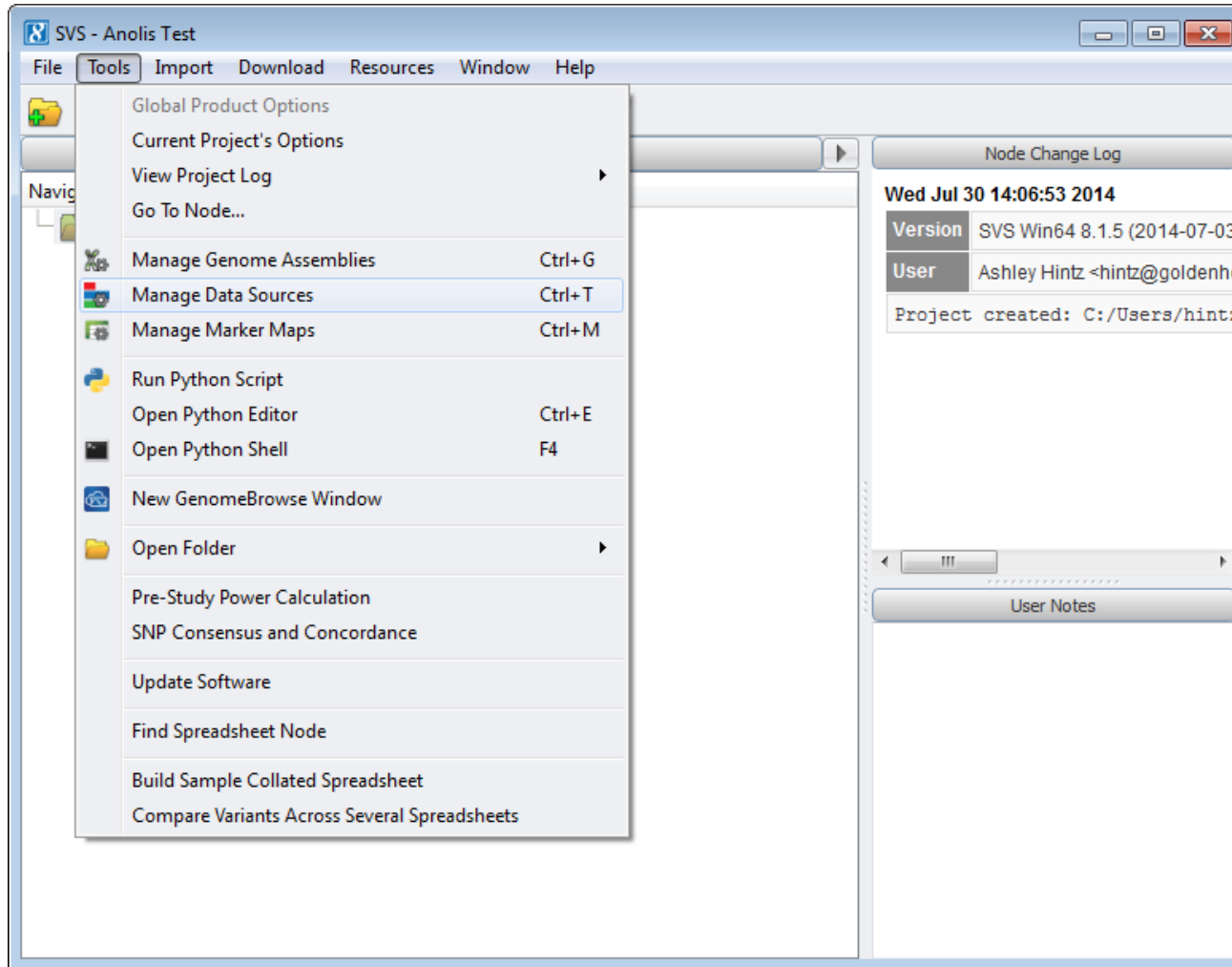


Photo courtesy R. Colin Blenis



Photo courtesy Robert Michniewicz

Build Custom Genome



Build Custom Genome



The screenshot displays the 'Data Source Library' window with a 'Convert Source Wizard' dialog box open. The wizard is in the 'Define Input' step, showing a list of data sources on the left and a 'Select Files' area on the right. The 'Add' button in the 'Select Files' area is circled in red. The 'Convert...' button in the bottom left of the main window is also circled in red.

Data Source Library

Locations: Local

Filter: *

Name

- 1kG Phase1 - Variant
- Cytobands 2009-06-1
- dbNSFP NS Function
- dbSNP Common 137
- NHLBI ESP6500SI-V2
- Reference Sequence
- RefSeq Genes 63, UC

Information

Local

Type Container

Convert Source Wizard

Convert Data Source

- 1 Define Input
- 2 Scan Input
- 3 Change Options
- 4 Convert

Select Files:

Select one or more files to Convert.

Files must be of the same type to be converted together.

Advanced Options

Build Custom Genome



The screenshot shows a software interface with a 'Data Source Library' window on the left and a 'Convert Source Wizard' dialog box in the foreground. The wizard is titled 'Convert Data Source' and has four steps: 1. Define Input, 2. Scan Input, 3. Change Options, and 4. Convert. The 'Define Input' step is active, showing a 'Select Files:' list with six files from 'Anolis_carolinensis.AnoCar2.0.74.dna.chromosome.1.fa' to '6.fa'. The 'Next >' button is circled in red. The 'Advanced Options' checkbox is checked, and a 'Help' button is visible at the bottom left of the wizard. The background window shows a file browser with a 'Local' folder selected, containing various genomic data sources like '1kG Phase1 - Variant' and 'RefSeq Genes 63, UC'.

Data Source Library

Locations: Local

Filter: *

Name

- 1kG Phase1 - Variant
- Cytobands 2009-06-1
- dbNSFP NS Function
- dbSNP Common 137
- NHLBI ESP6500SI-V2
- Reference Sequence
- RefSeq Genes 63, UC

Information

Local

Type Container

Convert Source Wizard

Convert Data Source

- Define Input
- Scan Input
- Change Options
- Convert

Select one or more files to Convert.

Files must be of the same type to be converted together.

Advanced Options

Help

Select Files:

- Anolis_carolinensis.AnoCar2.0.74.dna.chromosome.1.fa
- Anolis_carolinensis.AnoCar2.0.74.dna.chromosome.2.fa
- Anolis_carolinensis.AnoCar2.0.74.dna.chromosome.3.fa
- Anolis_carolinensis.AnoCar2.0.74.dna.chromosome.4.fa
- Anolis_carolinensis.AnoCar2.0.74.dna.chromosome.5.fa
- Anolis_carolinensis.AnoCar2.0.74.dna.chromosome.6.fa

Add Remove

< Back Next > Cancel

Convert... Export... Utilities Download Close Help

Build Custom Genome



The screenshot displays the 'Convert Source Wizard' window, which is part of a larger 'Data Source Library' application. The wizard is currently in the 'Scanning Genomic Coordinates Ranges' step, which is the second of four steps: 1. Define Input, 2. Scan Input, 3. Change Options, and 4. Convert. The progress bar shows 14% completion, with a 'Skip' button and a 'Next >' button. A red circle highlights the 'Next >' button. The text on the screen reads: 'Scanning the source to determine the segment names and lengths. This provides the details necessary to define a new genome assembly (build). About 20 Seconds remaining'. The 'Advanced Options' checkbox is checked, and a 'Help' button is visible. The background shows a file browser with various genomic data sources like '1kG Phase1 - Variant' and 'Cytobands 2009-06-1'.

Build Custom Genome



Convert Data Source

1 Define Input
2 Scan Input
3 **Change Options**
4 Convert

Select an existing genome assembly/build that matches your data.

When converting a reference sequence, you can also define a new assembly using the detected chromosome (segment) in the source.

Advanced Options

Help

Genome Assembly (Build): <<Create New>

New Genome Assembly/Build

Species: Anolis carolinensis

Common Name: Carolina Anole Lizard Taxonomy Id: [icon]

Build Name: AnoCar.2.0 GenBank Id: [icon]

Build Date: 2010-05-01 RefSeq Id: [icon]

Define Segments:

Use	Source	Renamed	Segment	Length	Aliases	Type
<input checked="" type="checkbox"/>	1	1	1	263920458		Autosome
<input checked="" type="checkbox"/>	2	2	2	199619895		Autosome
<input checked="" type="checkbox"/>	3	3	3	204416410		Autosome
<input checked="" type="checkbox"/>	4	4	4	156502444		Autosome
<input checked="" type="checkbox"/>	5	5	5	150641573		Autosome
<input checked="" type="checkbox"/>	6	6	6	80741955		Autosome

Rename segments by: Prefix [input] → [input] [Set Segment to Renamed]

< Back **Next >** Cancel

Build Custom Genome



Data Source Library

Locations: User Annotations

Filter: * (Any type) Homo sapiens (Human), GRCh37 hg19 (Feb 2009) Latest Plot Data

Name	Type	Size	Date	Url
1kG Phase1 - Variant Frequencies 3 with Major/Minor, GHI	Variant	479M	2012-06-02	C:/Users/hintz.Atalanta/AppData/Local/Golden...

Convert Source Wizard

Convert Data Source

1 Define Input
2 Scan Input
3 **Change Options**
4 Convert

Provide documentation for this source.

Advanced Options

Help

Source Definition

Name: Reference Sequence AnoCar2.0, Ensembl
Curated Date: 2014-07-30 4:27 PM
Curated By: Ashley Hintz
Series Name: ReferenceSequence-Ensembl
Version:

Fields

Note: Changing the name of fields may break the ability of a source to be plotted as a specialized track type.

	Orient	Type	Name
1	Locus	Byte Ar...	Data

Description | Credit | Notes | Meta

HTML description of source:

```
<p>Allele sequence track for Anolis carolinensis from Ensembl AnoCar2.0</p>
<p><a href="ftp://ftp.ensembl.org/pub/release-75/fasta/anolis_carolinensis/dna/">ftp://ftp.ensembl.org/pub/release-75/fasta/anolis_carolinensis/dna/</a></p>
```

Documentation Preview

Description

Allele sequence track for Anolis carolinensis from Ensembl AnoCar2.0

ftp://ftp.ensembl.org/pub/release-75/fasta/anolis_carolinensis/dna/

Source Credit

Curation Notes

Header Data

< Back **Next >** Cancel

Convert... Export... Utilities Download Close Help

Build Custom Genome



The screenshot shows the 'Data Source Library' window with a 'Convert Source Wizard' dialog box open. The wizard is titled 'Convert Data Source' and is ready to convert a data source named 'Reference Sequence AnoCar2.0, Ensembl'. The wizard has four steps: 1. Define Input, 2. Scan Input, 3. Change Options, and 4. Convert. The 'Convert' step is currently active. The wizard displays the following information:

- Input: ReferenceSequenceAnoCar2.0-Ensembl_AnoCar2.0_Anolis_carolinensis.tsf:1
- Total size: 292M
- Number of Fields: 1
- Detected track type: Allele Sequence
- Assembly: AnoCar2.0,Chromosome,Anolis carolinensis
- Coverage Computation: SequenceCoverage

The wizard also shows the file name and path:

File Name: ReferenceSequenceAnoCar2.0-Ensembl_AnoCar2.0_Anolis_carolinensis.tsf
Path: C:/Users/hintz.Atalanta/AppData/Local/Golden Helix/Common Data/Annotations

The 'Convert' button is highlighted with a red circle. The 'Advanced Options' checkbox is checked. The 'Add Path to Library' checkbox is unchecked. The 'Help' button is visible at the bottom left of the wizard. The 'Data Source Library' window in the background shows a list of data sources under the 'User Annotations' folder, including '1kG Phase1 - Variant', 'Cytobands 2009-06-30', 'dbNSFP NS Function', 'dbSNP 138, NCBI', 'dbSNP Common 137', 'NHLBI ESP6500SI-V2', 'Reference Sequence', and 'RefSeq Genes 63, UCSC'. The 'Information' section for 'User Annotations' is also visible, showing options like 'Type', 'Url', 'Include Sub-Folders', and 'Flatten Folder Structure'.

Build Custom Genome



The screenshot shows the 'Data Source Library' window with a list of data sources on the left and the 'Convert Source Wizard' dialog box in the foreground. The wizard is currently on the 'Convert' step (step 4 of 4). The source being converted is 'Anolis_carolinensis'. The progress bar shows 55% completion, and it indicates 'About 1 Minute remaining'. The 'Finish' button is circled in red.

Data Source Library

Locations: Local

Filter: *

Name

- 1kG Phase1 - Variant
- Cytobands 2009-06-1
- dbNSFP NS Function
- dbSNP Common 137
- NHLBI ESP6500SI-V2
- Reference Sequence
- RefSeq Genes 63, UC

Information

Local

Type: Container

Convert Source Wizard

Convert Data Source

- Define Input
- Scan Input
- Change Options
- Convert**

Convert the source.

Clicking Back will cancel the convert, but allow you to restart it.

Anolis_carolinensis

- ✓ Preparing source reader
- ▶ **Writing new file**
 - Computing SequenceCoverage on source

55%

About 1 Minute remaining

Advanced Options

Help

< Back Finish Cancel

Convert... Export... Utilities Download Close Help



GOLDEN HELIX SNP & VARIATION SUITE



[Demonstration]



Questions or more info:

- Email info@goldenhelix.com
- Request an evaluation of the software at www.goldenhelix.com

