

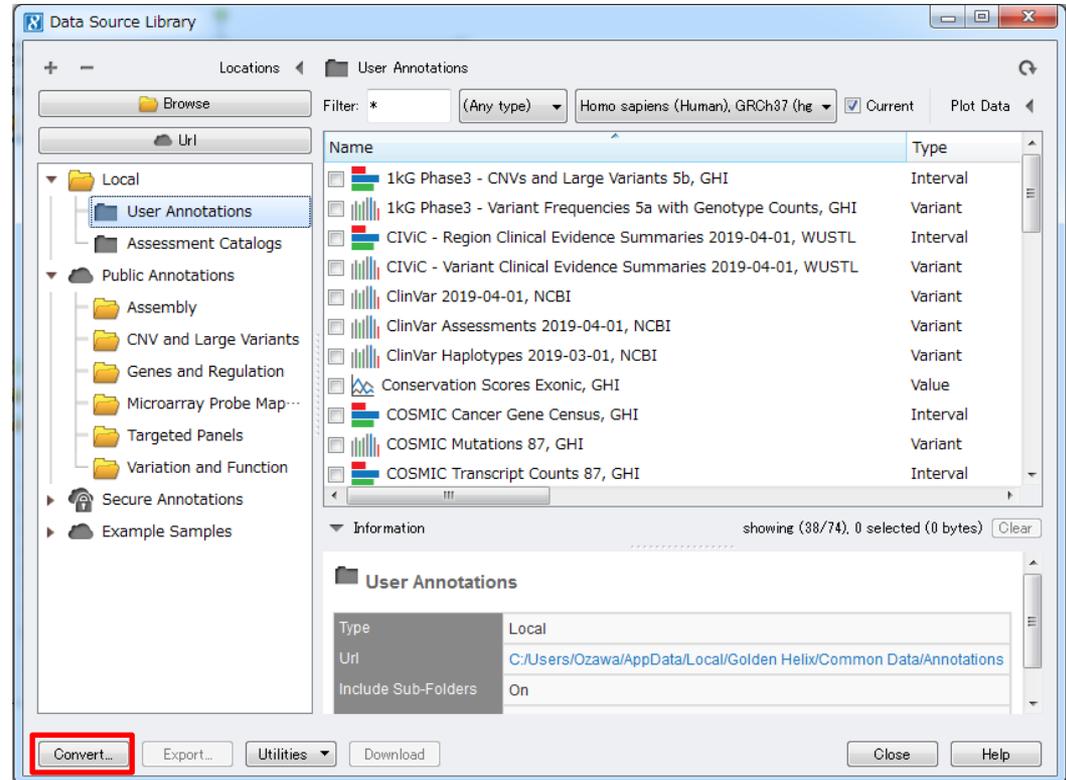
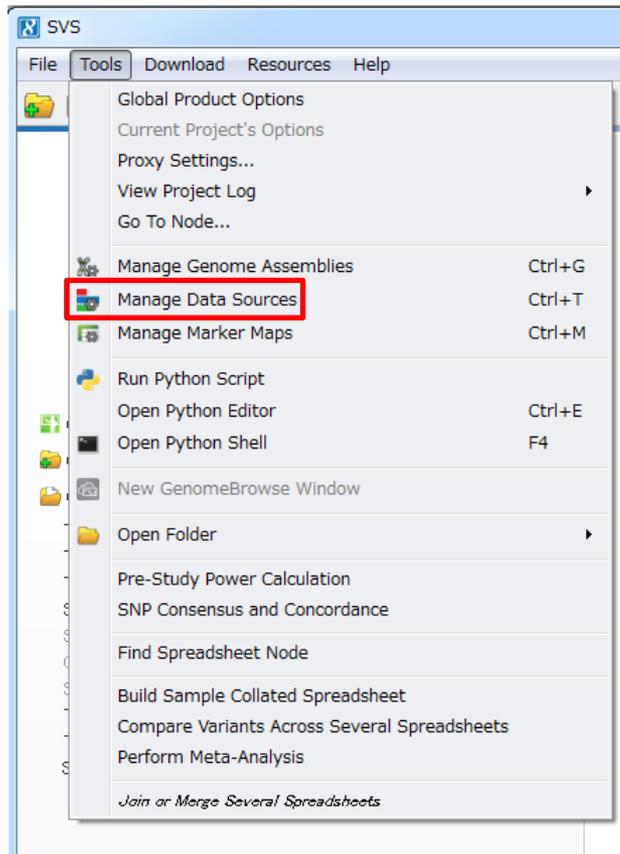
Genome Assemblies データの登録

フィルジェン株式会社 バイオインフォマティクス部
(biosupport@filgen.jp)

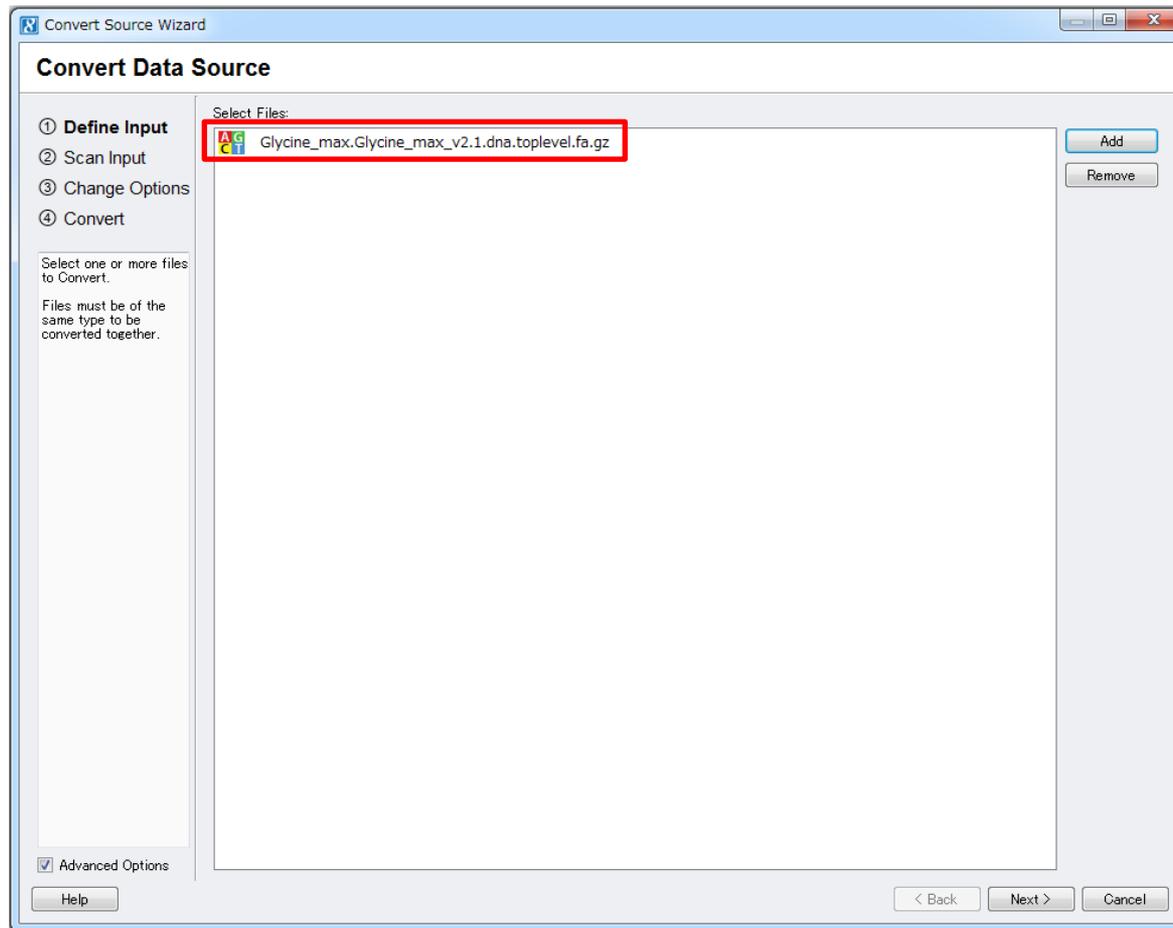
- SNP & Variation Suiteに新規ゲノム情報を登録し、各種解析に使用する場合、ゲノム配列データのFASTAファイル、およびオプションで遺伝子アノテーションデータのGTF（またはGFF）ファイルが必要です。
- ここでは、EnsemblPlants (<http://plants.ensembl.org/>) より取得した、[Glycine max v2.1 \(Jul 2018\)](#) のゲノム配列データと遺伝子アノテーションデータの登録方法を紹介します。

Dioscorea rotundata	FASTA (DNA)	FASTA (cDNA)	FASTA (CDS)	FASTA (ncRNA)	FASTA (protein)	EMBL	GenBank	MySQL(core)	TSV	GTF	GFF3
Galdieria sulphuraria	FASTA (DNA)	FASTA (cDNA)	FASTA (CDS)	FASTA (ncRNA)	FASTA (protein)	EMBL	GenBank	MySQL(core)	TSV	GTF	GFF3
Glycine max	FASTA (DNA)	FASTA (cDNA)	FASTA (CDS)	FASTA (ncRNA)	FASTA (protein)	EMBL	GenBank	MySQL(core) MySQL(funcgen)	TSV	GTF	GFF3
Gossypium raimondii	FASTA (DNA)	FASTA (cDNA)	FASTA (CDS)	FASTA (ncRNA)	FASTA (protein)	EMBL	GenBank	MySQL(core) MySQL(funcgen)	TSV	GTF	GFF3

1. データベースのダウンロードページより、FASTA (DNA)ページ内の任意の*.fa.gzファイル（例：Glycine_max.Glycine_max_v2.1.dna.toplevel.fa.gzなど）をダウンロード
2. 同じくGTFページ内の任意の*.gtf.gzファイル（例：Glycine_max.Glycine_max_v2.1.43.gtf.gzなど）をダウンロード
* GTFの代わりに、GFFファイルも使用可能



1. SNP & Variation Suiteを起動し、メニューのTools -> Manage Data Sourcesをクリック
2. Data Source Library画面より、Convertをクリック



3. ① Define Input画面でAddをクリックし、ファイル選択画面でファイルタイプをFASTA Filesに指定したうえで、ゲノム配列データのFASTAファイルを選択
4. Nextをクリックするとファイルのスキャンが開始されるので、スキャンが終了したらNextをクリック

Convert Data Source

Genome Assembly (Build): <<Create New>>

New Genome Assembly/Build

Species: Glycine max

Common Name: Soybean Taxonomy Id: [icon]

Build Name: Glycine_max_v2.1 GenBank Id: [icon]

Build Date: 2018-07-01 RefSeq Id: [icon]

Define Segments:

Use	Source	Renamed	Segment	Length	Aliases	Type	Visibility
<input checked="" type="checkbox"/>	18 dnachrom...	18	18	58018742		Autosome	Always
<input checked="" type="checkbox"/>	1 dnachrom...	1	1	56831624		Autosome	Always
<input checked="" type="checkbox"/>	4 dnachrom...	4	4	52389146		Autosome	Always
<input checked="" type="checkbox"/>	15 dnachrom...	15	15	51756343		Autosome	Always
<input checked="" type="checkbox"/>	10 dnachrom...	10	10	51566898		Autosome	Always
<input checked="" type="checkbox"/>	6 dnachrom...	6	6	51416486		Autosome	Always
<input checked="" type="checkbox"/>	19 dnachrom...	19	19	50746916		Autosome	Always
<input checked="" type="checkbox"/>	9 dnachrom...	9	9	50189764		Autosome	Always
<input checked="" type="checkbox"/>	14 dnachrom...	14	14	49042192		Autosome	Always
<input checked="" type="checkbox"/>	2 dnachrom...	2	2	48577505		Autosome	Always
<input checked="" type="checkbox"/>	20 dnachrom...	20	20	47904181		Autosome	Always
<input checked="" type="checkbox"/>	8 dnachrom...	8	8	47837940		Autosome	Always
<input checked="" type="checkbox"/>	13 dnachrom...	13	13	45874162		Autosome	Always
<input checked="" type="checkbox"/>	3 dnachrom...	3	3	45779781		Autosome	Always

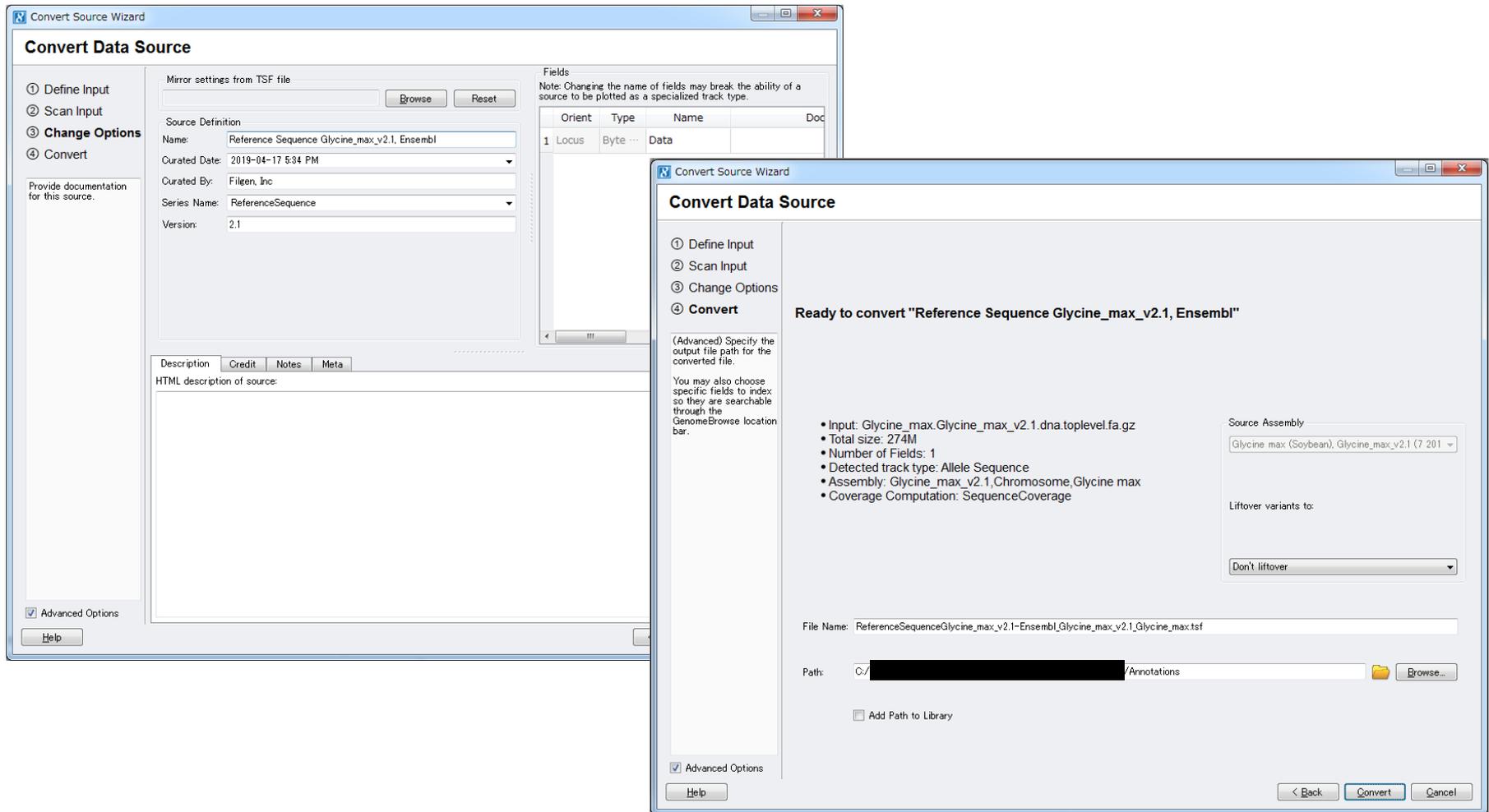
Rename segments by: RegExp (*). dna(*) → %1

Set Segment to Renamed

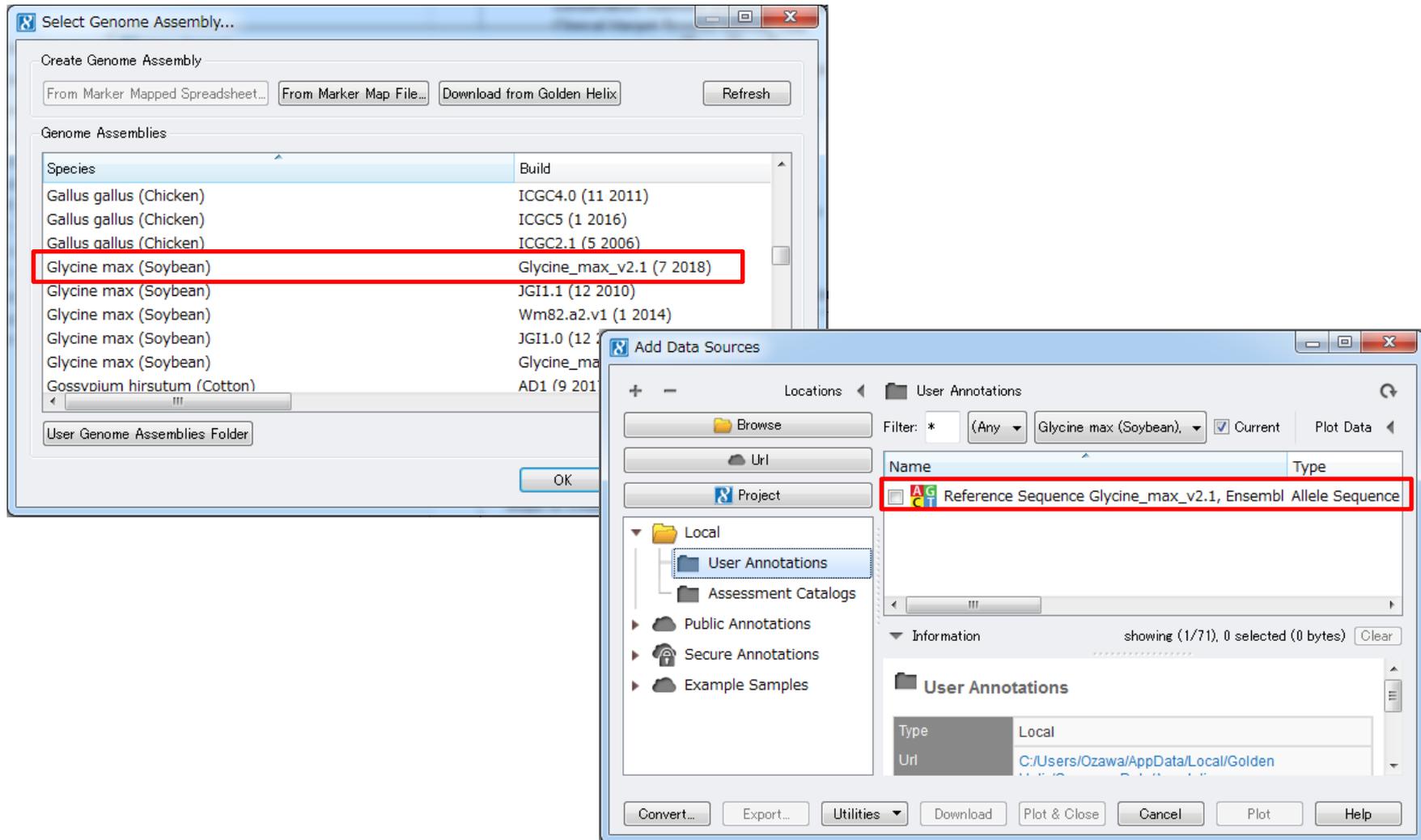
5. 必要に応じて、Genome Assemblyの生物種情報やBuild情報などを入力

6. FASTAファイルに含まれる配列名を染色体番号に編集するため、Rename segmentを「RegExp」と指定し、1つ目の入力スペースに「(*) dna(*)」、2つ目に「%1」と入力

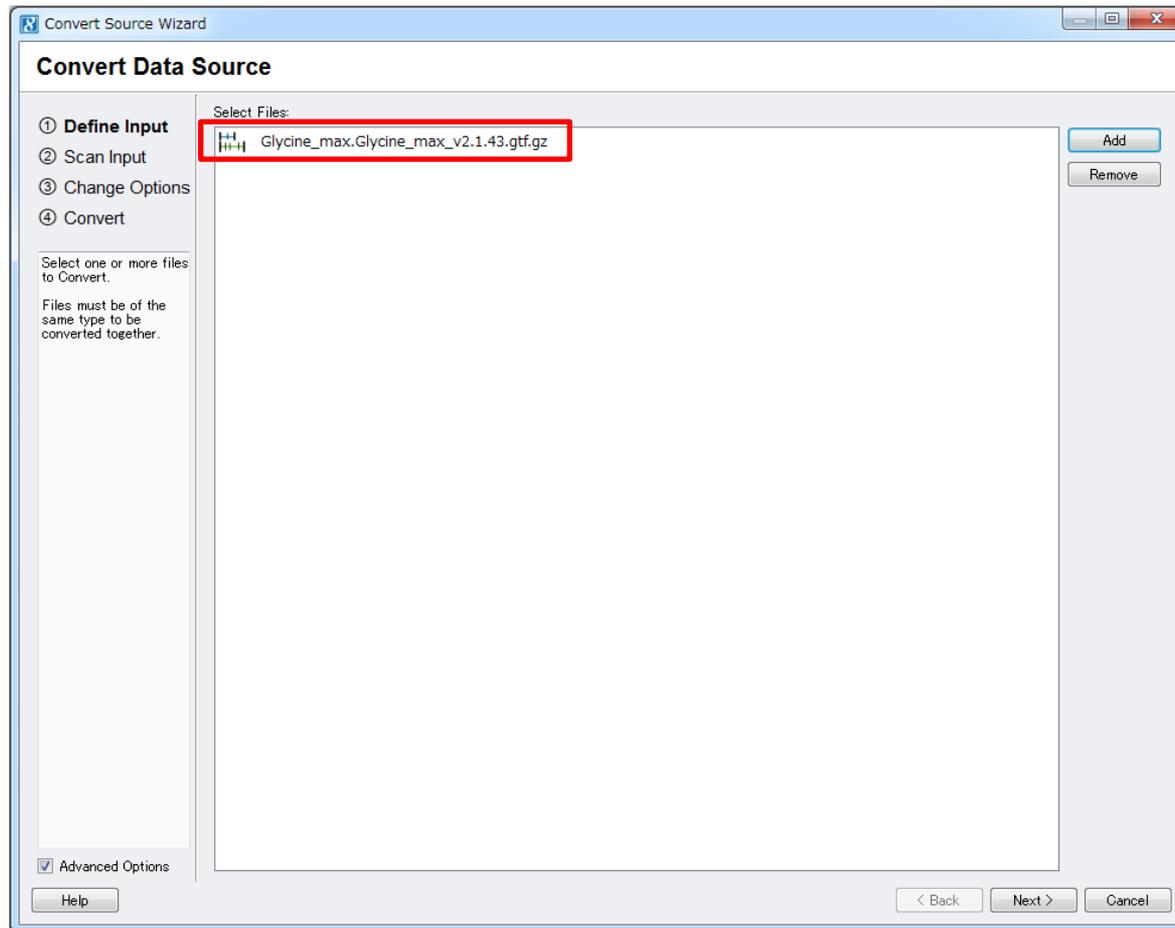
7. Renamedフィールドで、配列名が染色体番号に正しく編集されていたら、Set Segment to Renamedをクリックして、編集結果をSegmentフィールドに反映させる



8. ③ Change Options画面で、必要に応じてデータの詳細情報を入力
9. ④ Convert画面で、ファイル名や保存先などを確認した後Convertをクリック
10. Convertが終了したらFinishをクリック



11. Convertが終了すると、Genome Assemblyで登録したゲノムを選択できるようになり、またData Source LibraryのUser Annotationsに、インポートしたゲノム配列トラックが保存される



1. FASTAインポートと同様に、Convert Source Wizardの① Define Input画面でAddをクリックし、ファイル選択画面でファイルタイプをGTF Filesに指定したうえで、遺伝子アノテーションデータのGTFファイルを選択
2. Nextをクリックするとファイルのスキャンが開始されるので、スキャンが終了したらNextをクリック

Convert Data Source

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

You can rename, drop, reorder and change the type of fields.
Note that when data fails to be coerced into a specified type, it will be set as missing.

Desired Plot Type: (Automatically Detect) ⓘ
Detected Plot Type: Gene
Edit Output Fields:

Use	Input	Name	Type	Orient
<input checked="" type="checkbox"/>	Gene Name	Gene Name	String	Locus
<input checked="" type="checkbox"/>	Transcript Na...	Transcript Name	String	Locus
<input checked="" type="checkbox"/>	CDS Start	CDS Start	Int	Locus
<input checked="" type="checkbox"/>	CDS Stop	CDS Stop	Int	Locus
<input checked="" type="checkbox"/>	Exon Starts	Exon Starts	Int Array	s
<input checked="" type="checkbox"/>	Exon Stops	Exon Stops	Int Array	s
<input checked="" type="checkbox"/>	Strand	Strand	String	Locus
<input checked="" type="checkbox"/>	Source	Source	String	Locus
<input checked="" type="checkbox"/>	exon_id	exon_id	String	Locus
<input checked="" type="checkbox"/>	exon_number	exon_number	String	Locus
<input checked="" type="checkbox"/>	gene_biotype	gene_biotype	String	Locus

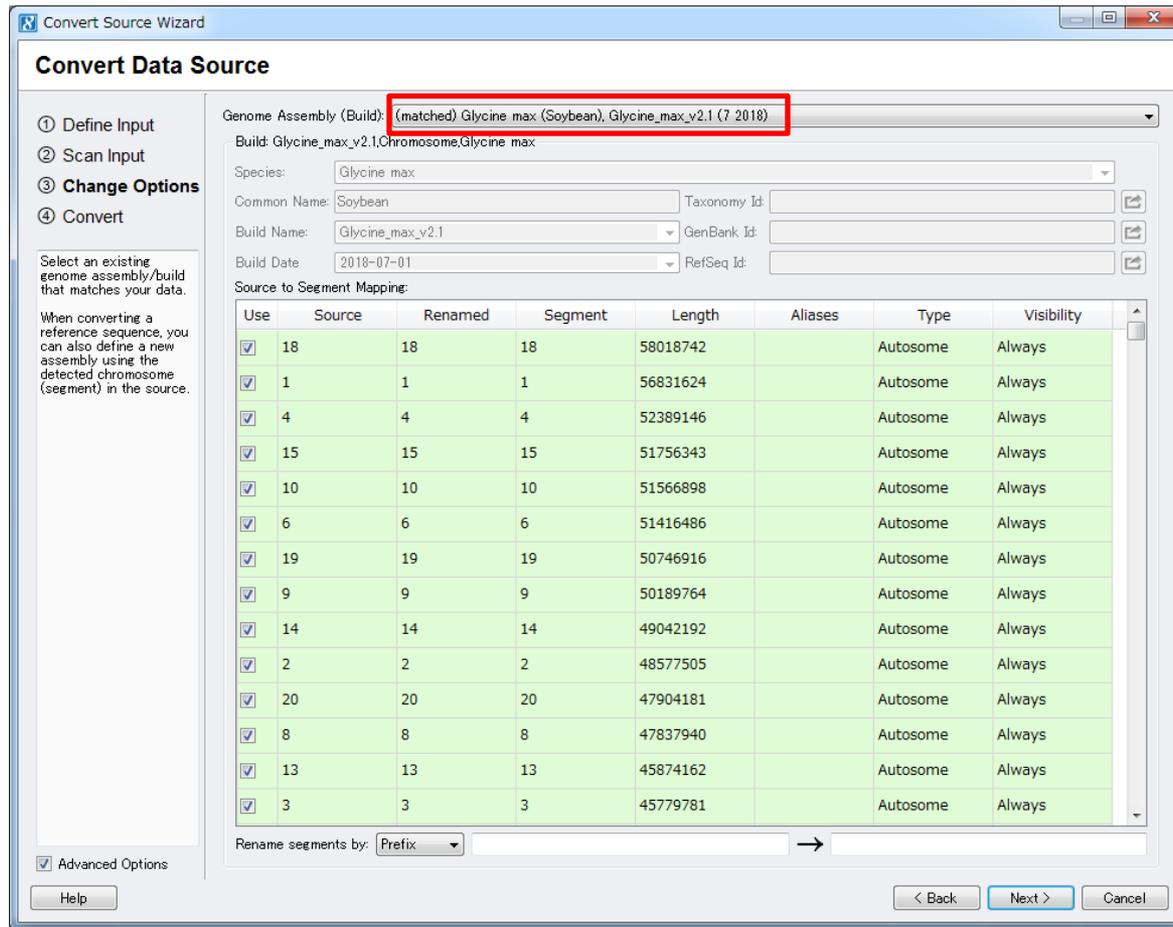
Preview: 1000 features read into preview ([Read More](#))

	Chr	Start	Stop	Gene Name	Transcript Name	CDS Start	CDS Stop	Exon Starts	
1	18	2448	40092	GLYMA_18G0...	KRG97315	2716	39647	2447,3738,38...	27
2	18	47154	49570	GLYMA_18G0...	KRG97319	47470	48505	47153	49
3	18	47471	48505	GLYMA_18G0...	KRG97316	47470	48505	47470	48
4	18	47471	48505	GLYMA_18G0...	KRG97317	47470	48505	47470	48
5	18	47471	48505	GLYMA_18G0...	KRG97318	47470	48505	47470	48
6	18	52031	55923	GLYMA_18G0...	KRG97320	52162	55306	52030,52393,...	52
7	18	58631	61226	GLYMA_18G0...	KRG97321	58760	61129	58630,59535,...	59
8	18	83396	88423	GLYMA_18G0...	KRG97322	83980	88225	83395,87987	86
9	18	99110	101506	GLYMA_18G0...	KRG97323	99474	101151	99109,99870,...	99
10	18	103443	105905	GLYMA_18G0...	KRG97324	104066	105458	103442	10
11	18	106361	110715	GLYMA_18G0...	KRG97325	106892	110289	106360,1077...	10

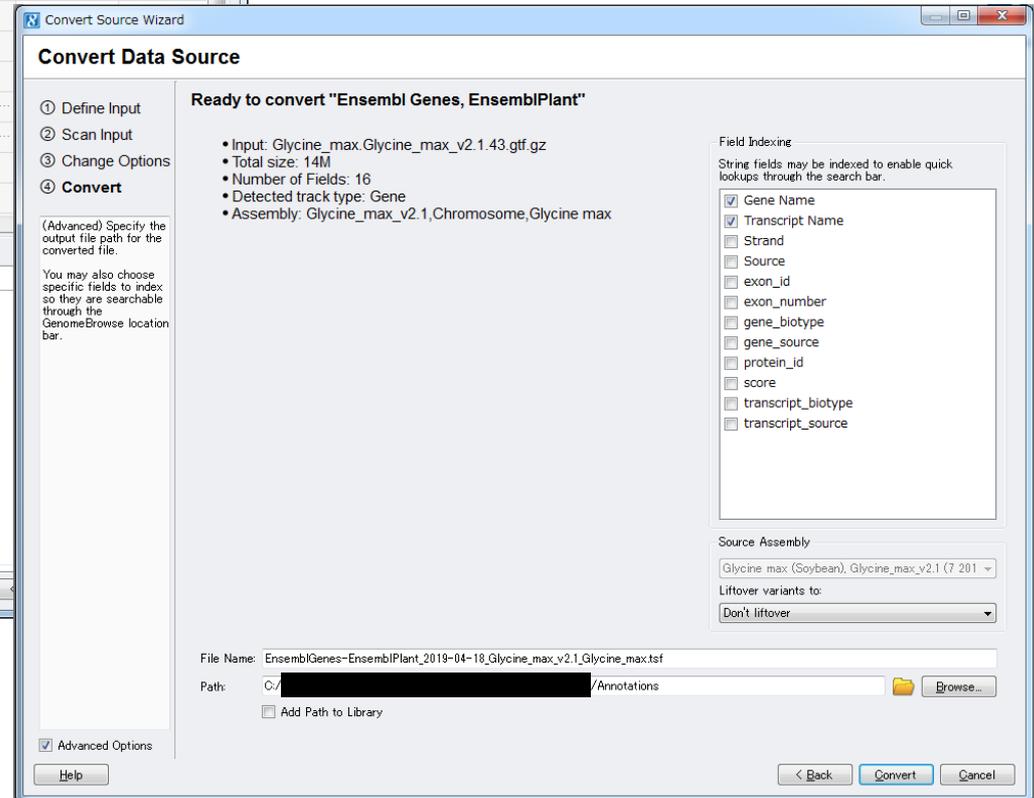
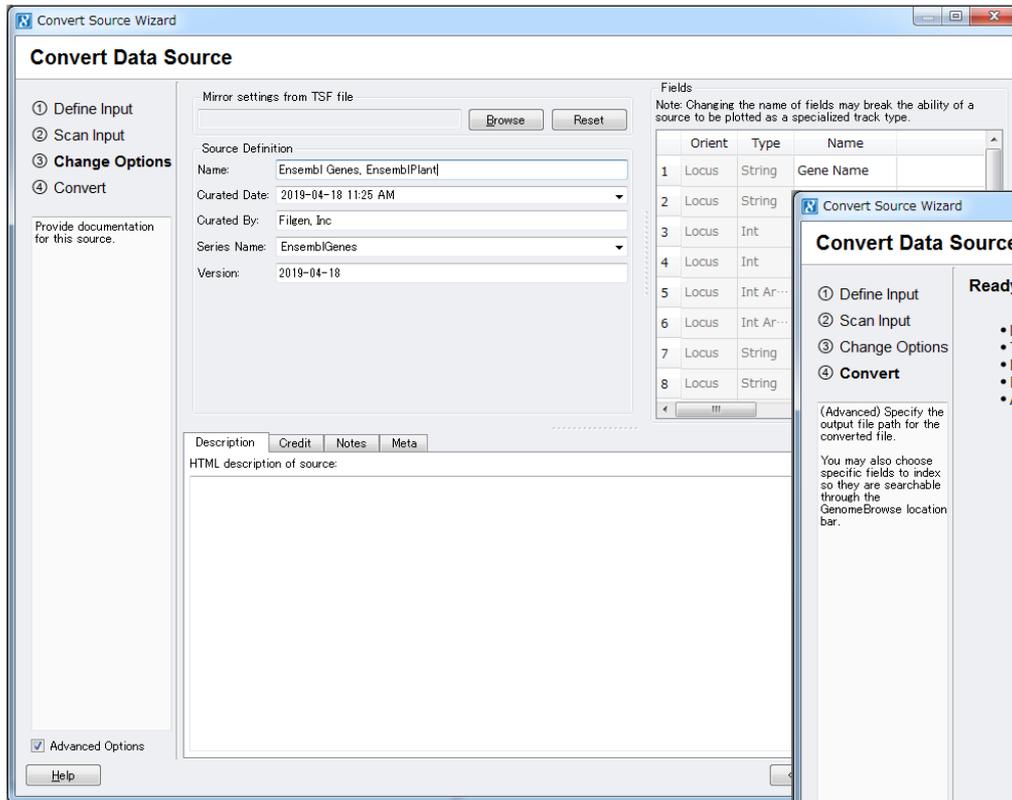
Advanced Options

Help < Back Next > Cancel

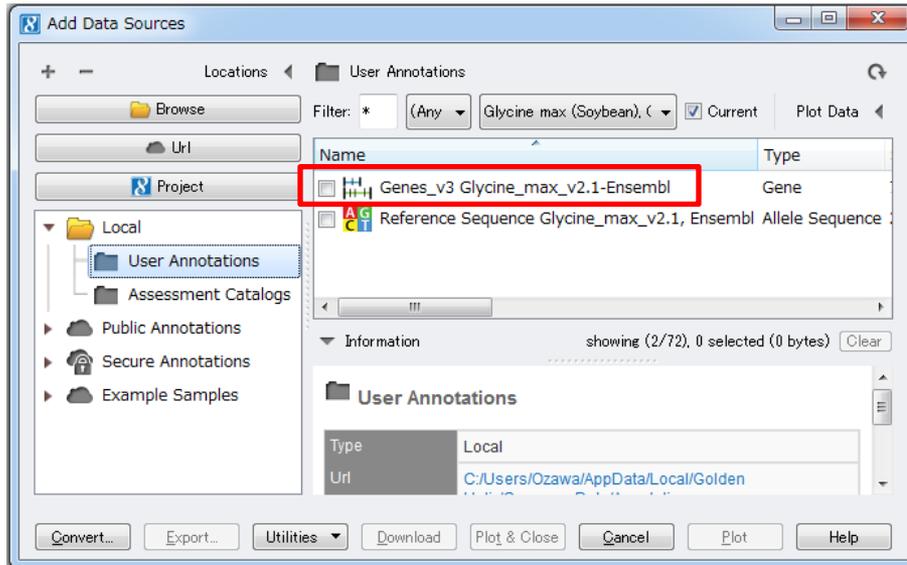
3. スキャンが終了すると、ファイルに含まれているデータフィールドなどが表示されるので、内容を確認してNextをクリック



4. 先にFASTAファイルからインポートしておいたGenome Assembly情報が、Genome Assembly (Build)で選択されていることを確認して、Nextをクリック

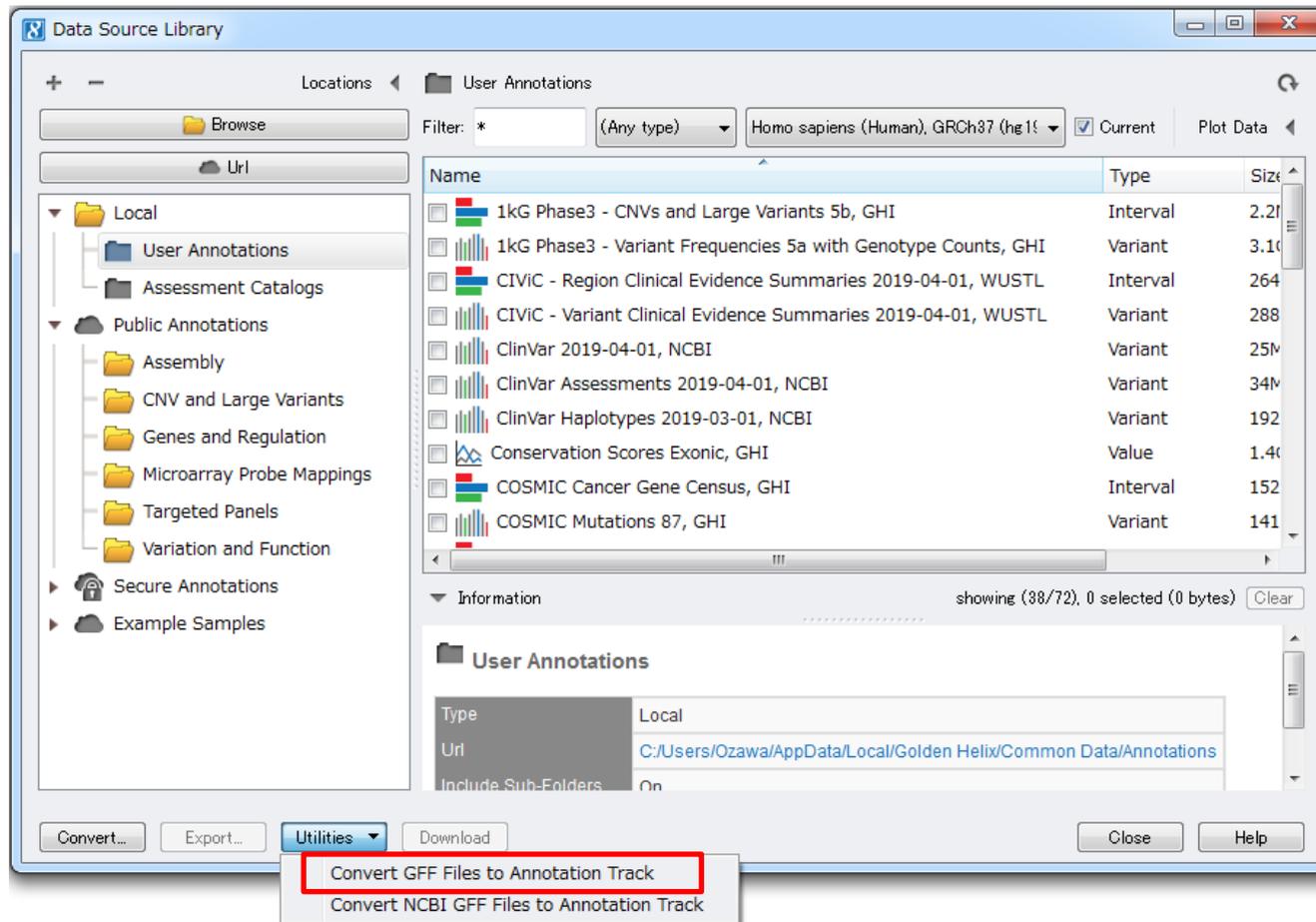


5. ③ Change Options画面で、必要に応じてデータの詳細情報を入力
6. ④ Convert画面で、ファイル名や保存先などを確認した後Convertをクリック
7. Convertが終了したらFinishをクリック

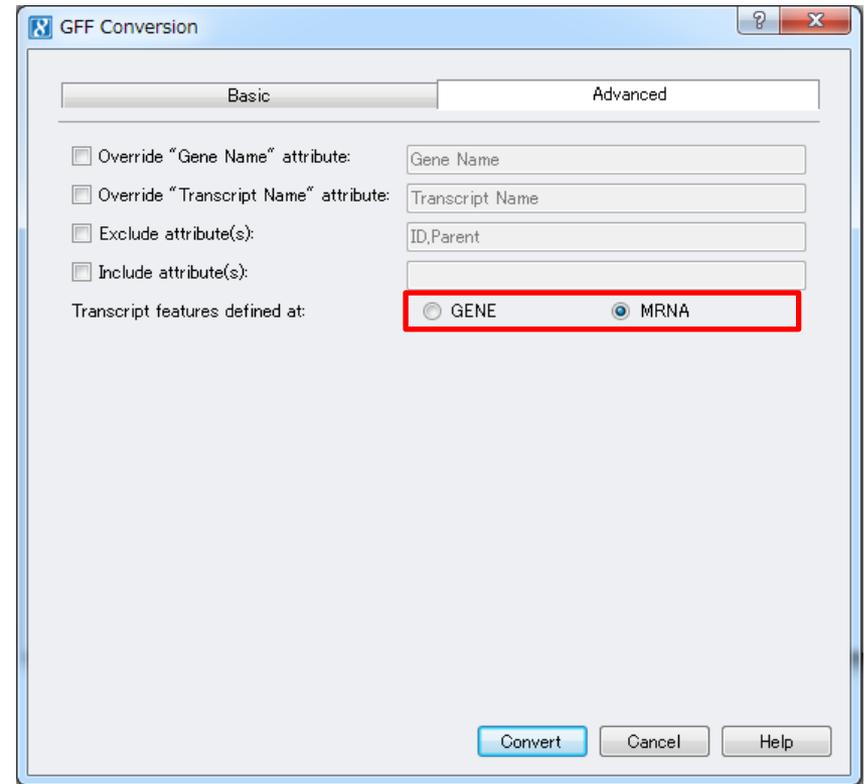
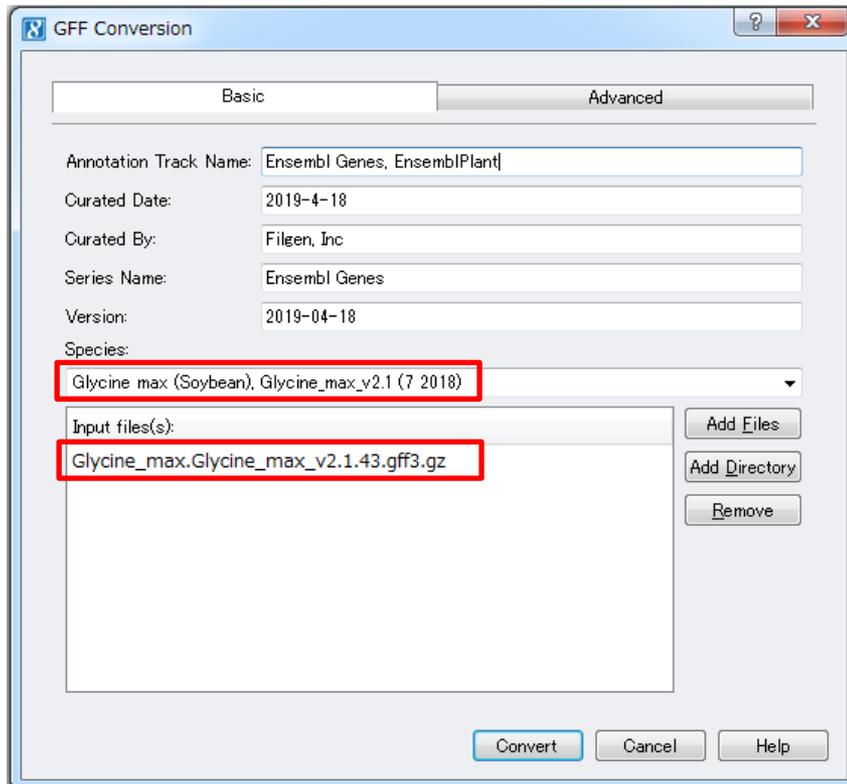


8. Convertが終了すると、Data Source LibraryのUser Annotationsに、インポートした遺伝子アノテーショントラックが保存される
9. ゲノム配列トラックと遺伝子アノテーショントラックの両方とも適切にインポートされていれば、両トラックをゲノムブラウザーに表示ができる

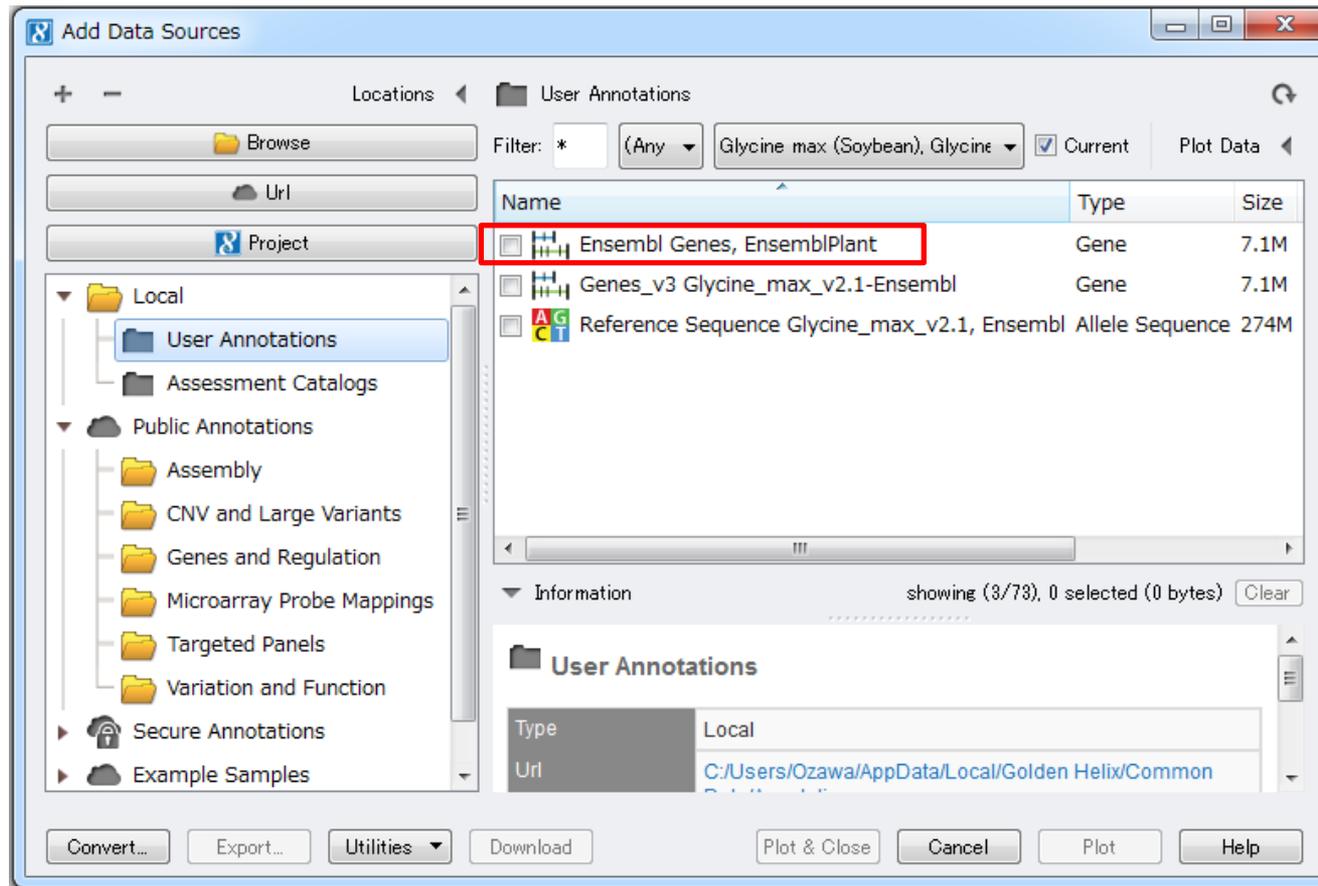




1. Data Source Library画面より、Utilities -> Convert GFF Files to Annotation Trackをクリック



2. 「Species:」にて、先にFASTAファイルからインポートしておいたGenome Assemblyを選択し、「Input files(s):」にGFFファイル、さらに必要に応じてデータの詳細情報を入力
3. EnsemblやPhytozome (<https://phytozome.jgi.doe.gov/pz/>)よりダウンロードしたGFF3ファイルの場合、Advancedタブの「Transcript features defined at:」で「MRNA」を選択し、Convertをクリック



4. Convertが終了すると、Data Source LibraryのUser Annotationsに、インポートした遺伝子アノテーショントラックが保存される

お問い合わせ先：フィルジェン株式会社

TEL: 052-624-4388 (9:00～18 : 00)

FAX: 052-624-4389

E-mail: biosupport@filgen.jp