



第27回 リファレンスマッピングと SNP 検出 (その3)



NGS リードをリファレンス配列にマッピングし、SNP を検出するための一般的なワークフローについ てご紹介しています。

<u>第26回 リファレンスマッピングと SNP 検出(その1)</u> <u>第27回 リファレンスマッピングと SNP 検出(その2)</u>

3回目となる今回は、コンティグから SNP を検出する方法についてです。

Geneious Prime では Variant Finder 機能を使用することで、マッピングされたデータから SNP を検出することができます。コンティグドキュメントを選択し、Annotate and Predict → Find Variations/SNPs と進みます。

一番上にある Find Variants パネルにあるオプションで、変異の頻度や p 値などをしきい値として SNPをコールする時のパラメータを設定し、シークエンスエラーに起因する不一致をフィルタリングで 除外することができます。これらの設定に関する詳細な情報が必要な場合は?ボタンをクリックする か、各オプションにマウスオーバーしてください。

Analyze effect of variants on translations オプションにチェックし、Default Genetic Code を設定することで、リファレンス配列の CDS アノテーションを使用して、マッピングされたリードのコード配列を決定し、検出された SNP がアミノ酸配列に変化をもたらすかどうかを計算することができます。チュートリアルデータの場合は Bacterial に変更し、その他の設定はそのままで OK ボタンをクリックすると解析が開始されます。

	Find Variations/SNPs
-ind Variants	
Minimu	m Coverage: 1 🗇
🗹 Minimum Variar	nt Frequency: 0.25 🗇
🗹 Maximum Variant P-val	ue: 10 ^{- 6} 💭 (0.0001% to see variant by chance) ?
🗹 Minimum Strand-Bias P-	value: 10 ^{- 5} 🗘 when exceeding 65 🗘 % bias ?
	Find Variants Inside & Outside CDS 🛛 🗸
	In selected region only
Analyze Effects on Translation	S
Analyze Effects on Translation Analyze effect of variants o Default Genetic Code:	s on translations
Analyze Effects on Translation Analyze effect of variants o Default Genetic Code: Calculate Variant P-values	s on translations Bacterial V
Analyze Effects on Translation Analyze effect of variants o Default Genetic Code: Calculate Variant P-values Assumed quality of bases with	s on translations Bacterial Nout quality: 20
Analyze Effects on Translation Analyze effect of variants Default Genetic Code: Calculate Variant P-values Assumed quality of bases with P-value calculation method:	s on translations Bacterial Nout quality: 20 (99.0% correct) Approximate
 Analyze Effects on Translation Analyze effect of variants of Default Genetic Code: Calculate Variant P-values Assumed quality of bases with P-value calculation method: Homopolymer quality reduced 	s on translations Bacterial Nout quality: 20

解析が終了すると、コンティグに Variants のアノテーショントラックが追加されます(チュートリアル データの場合は Variants: yghJ paired Illumina reads (trimmed))。**Save** をクリックする と、オリジナルの配列に変更を適用するか確認されますので、**Yes** を選択して、アノテーショントラッ クを元のリファレンス配列に保存します。

Contig View Annotations	Lengths Graph Insert Sizes Text View Lineage Info
$\leftarrow \rightarrow \bigcirc$ Extract @ R.C.	🕅 Translate 🗠 Add Annotation 🥢 Allow Editing 🖒 Annotate & Predict 🕞 Save
1 200 400 600	800 1,000 1,000 1,000 1,000 1,000 2,000 2,000 2,000 2,000 2,000 3,000 3,000 3,000 3,000 3,000 4,000 4,000 4,000
	130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280
Consensus	νε ασσστετττσε ε σαλασταλλες τα πτε σλελές λλες ε σα σε σε στα σε στα στα παραγικά το παραγικά το παραγικά το παραγικά το παραγικά το παραγικά το παραγικό το παραγικ
Couprage 1321	
coverage ol	
	3196.847 3196.837 3196.837 3196.837 3196.817 3196.807 3196.797 3196.787 3196.777 3196.767 3196.757 3196.747 3196.737 3196.727 3196.717 3196.707 3196.707
De 1. ygh] CDS (divergent ref	LAGGETETTER E GRANT GAMARE TRATE CARCACCARCECCEGARE CORE CARCERCERCE CARCECCARCECCEGARE CORE CORE CORE CORE CORE CORE CORE CO
	yghi CDS >>
	>> yghl gene >>
 Coverage: ygh] paired Illumi. 	31->38
- Varianter unbl paired Illumin	Apportation track
 variants: ygrij paireo iliumin 	Annotation track 4,578 bp & 3 gaps
C+ F=0 == 65. SRR513053.420	
De FID == 107 SPR513053.254	IR ADDRUCT THE CORPARISON OF THE ANALOUS AND CONCERNMENT OF THE ADDRUCT OF THE AD
C. FOD = 68. SRR513053.261	
C. FRO == 69. SRR513053.144	
Contraction 100 State 100	
De FID = 71 SRS13053.23	
C. FID - 72. SRR513053.129	
C+ FIID == 109. SRR513053.11	KEAGGGTETTTEEEGGAAGTGAAAECTGATEEAAEAEEGGAAECEGAEGEEGGAEGEEGGAEGEEGGAEGEEGGAEGEEGGAAEEGGAAEEGGAAEECGAEGEEGGAEGE
De Feo - 73. SRR513053.773	ICAGGOTETTTECEGGAAGTAAAECTGATECAAEACCGGAGECGGTGAACCGBEOCCTGACC
Co FID - 75 SRS13053.264	IL AND IL THE COMPACTION AND FOR THE ANAL COMPACTION OF THE CARLES OF TH
De FID = 76. SRR513053.279	
Co FILO == 77. SRR513053.199	TEREGORANGE CAAR ARE ARE ARE ARE ARE ARE CONSIDERED AND CONSIDER
De FID - 78. SRR513053.126	ECIDANGTOANARETOATECAARAECEGGREGCEGGEGCETGTECTADEAAAA
0 FID - 90 SRK513053.7/1	DIGRANG FOR AND E FOR THE CARACTERISANCE CONTROLS OF TARGET AND
De FID = 81, SRR513053.239	
De FID =- 82. SRR513053.264	EGAAGTGANACCTGATCCAACACCAGAACACCTGATCCTAGGCCTGAGCCTGAGCCTGAGCCTGAGCCTGAGCCGCAGACCCGAGACCGCAGACCGAGACCGCGCGGAGCCGGAGCCGGAGCCGGAGCCGGGGGG
De Fill =- 83. SRR513053.179	RETURARCETORTCEARCRECAGARCEGORGECORECTORSECTORSECTORSECTORSECTORSECTORSECTORSECTORSECTORTCETREGRAME
C 100 - 84. SRR513053.148	TGRARACCTEGATE CARACTEGAGE CONCECTOR OF THE ACCOUNT OF A CONCECTOR ACCOUNT OF THE
De FID == 86. SRR513053.230	EXAMPLETER TRANSPORTED AND TRA
De Fill = 87. SRR513053.116	DATECAACACCAGACCGGACCGACCGACCGACCGACCGACCGA

ご注意! 変異検出や発現解析結果を Save する時、元の(マッピング時にリファレンスとして用いた) 配列に変更内容(解析結果)を保存するかどうかを確認するダイアログが表示されます。ここで No を 選択した場合、この解析結果はこれ以降永久に、リファレンス配列に適用できなくなりますのでご注意 ください(再度の適用のためにはマッピングからのやり直しが必要になります)。変異をサンプルごと に比較する場合や、発現差解析を行う場合には、変異や発現量のアノテーションが、同じドキュメント (通常は元のリファレンス配列)内に存在する必要があります。

コンティグをスクロールして、SNP のアノテーション(黄色の縦棒)を含む位置まで移動します。アノテ ーションにマウスオーバーすると、その SNP に関する情報を含むポップアップウィンドウが表示され ます。これには塩基変化、バリアント頻度、SNP タイプ、タンパク質や CDS の変化に関する情報な どが含まれています。

Name:
Type: Polymorphism
Track: Variants: yghJ paired Illumina reads properties
Length: 3
Interval: <u>193 - 195</u> (3,196,782 - 3,196,784)
Bases: GAG
Change: -GAG
Reference Nucleotide(s): GAG
Coverage: 48
Variant Nucleotide(s):
Variant Frequency: 100.0%
Variant Raw Frequency: 48
Strand-Bias: 100.0%
Variant P-Value (approximate): 2.5 × 10 ⁻¹⁰⁶
Strand-Bias >50% P-value: 7.1 × 10 ⁻¹⁵
Strand-Bias >65% P-value: 2.1 × 10 ⁻⁹
Polymorphism Type: Deletion
CDS Codon Number: 65
CDS Position: 193
CDS Position Within Codon: 1
gene: yghJ
product: inner membrane lipoprotein
protein_id: YP_002388453.1
locus_tag: ECIAI1_3114
note: Evidence 2b : Function of strongly homologous gene;PubMedId : 1644747; Product type lp : lipoprotein
CDS: yghJ CDS
Codon Change: GAG ->
Amino Acid Change: E ->
Protein Effect: Deletion
Average Quality: 23

これらの情報をテーブルで表示するには、まずシークエンスビューアの上にある Annotations タブ をクリックします。すると配列上のすべてのアノテーションがテーブル表示されますので、Type をク リックし、Polymorphismを選択すると、多型アノテーションのみが表示されます。多型の種類、変異 頻度、アミノ酸/コドン/塩基の変化など、関連するカラムが自動的に表示されます。追加のカラムを 表示したり、既存のカラムを削除したりするには、Columns をクリックすることで、必要なカラムを 追加/削除することができます。

Contig View	Annotations	Lengths	Graph	Insert Sizes	Text View Li	neage Info							
⇒ Type: Polyr	morphism 强 Ti	rack: All ITt	Columns	🖞 Export t	able 🖆 Edit Annota	tion ⊖Extra	ct 🖉 Translate	Save					
Name	Minimum v	Maximum	Length	# Intervals	Amino Acid Change	CDS Position	Change	Codon Change	Coverage	Polymorphism Type	Protein Effect	Variant Frequency	Variant P-Value (approxi
ACA	1,942	1,944	3	1	L->T	1,942	CTG -> ACA	CTG -> ACA	110 -> 113	Substitution	Substitution	100.0%	0.0
т	1,941	1,941	1	1		1,941	G -> T	GTG -> GTT	109	SNP (transversion)	None	100.0%	0.0
т	1,935	1,935	1	1		1,935	G -> T	GGG -> GGT	114	SNP (transversion)	None	100.0%	0.0
CGT	1,927	1,929	3	1	K -> R	1,927	AAA -> CGT	AAA -> CGT	108 -> 109	Substitution	Substitution	100.0%	0.0
т	1,923	1,923	1	1		1,923	C -> T	TTC -> TTT	110	SNP (transition)	None	100.0%	0.0
с	1,917	1,917	1	1		1,917	G -> C	GTG -> GTC	100	SNP (transversion)	None	100.0%	0.0
т	1,914	1,914	1	1		1,914	$G \rightarrow T$	ACG -> ACT	101	SNP (transversion)	None	100.0%	0.0
r	1,911	1,911	1	1	E -> D	1,911	G -> T	GAG -> GAT	103	SNP (transversion)	Substitution	100.0%	0.0
c	1,902	1,902	1	1		1,902	T -> C	ACT -> ACC	100	SNP (transition)	None	100.0%	0.0
с	1,899	1,899	1	1		1,899	$T \Rightarrow C$	GGT -> GGC	97	SNP (transition)	None	100.0%	7.9E-321
A	1,896	1,896	1	1		1,896	G -> A	GTG -> GTA	102	SNP (transition)	None	100.0%	0.0
с	1,887	1,887	1	1		1,887	T -> C	AGT -> AGC	105	SNP (transition)	None	99.0%	0.0
5	1,882	1,882	1	1	S -> A	1,882	T -> G	TCC -> GCC	103	SNP (transversion)	Substitution	100.0%	0.0
5	1,878	1,878	1	1		1,878	A -> G	GCA -> GCG	102	SNP (transition)	None	100.0%	0.0
c	1,875	1,875	1	1		1,875	T -> C	GAT -> GAC	104	SNP (transition)	None	100.0%	0.0
AA.	1,867	1,868	2	1	L-> K	1,867	TT -> AA	TTG -> AAG	98 -> 99	Substitution	Substitution	100.0%	0.0
AAA	1,861	1,863	3	1	R -> K	1,861	CGC -> AAA	CGC -> AAA	98 -> 99	Substitution	Substitution	100.0%	0.0
GAC	1,855	1,857	2	2	N -> D	1,855	AAT -> GAC	AAT -> GAC	100 -> 102	Substitution	Substitution	100.0%	0.0
c	1,854	1,854	1	1		1,854	A -> C	TCA -> TCC	102	SNP (transversion)	None	100.0%	0.0
r	1,735	1,735	1	1		1,735	C -> T	CTG -> TTG	95	SNP (transition)	None	98.9%	2.4E-318
r	1,734	1,734	1	1		1,734	C -> T	ATC -> ATT	95	SNP (transition)	None	100.0%	1.0E-323
Α.	1,698	1,698	1	1		1,698	G -> A	AAG -> AAA	101	SNP (transition)	None	100.0%	0.0
G	1,692	1,692	1	1	D -> E	1,692	T -> G	GAT -> GAG	99	SNP (transversion)	Substitution	100.0%	0.0
r	1,689	1,689	1	1		1,689	C -> T	GGC -> GGT	100	SNP (transition)	None	100.0%	0.0
c	1,686	1,686	1	1		1,686	G -> C	GTG -> GTC	101	SNP (transversion)	None	100.0%	0.0
c	1,557	1,557	1	1		1,557	T -> C	TTT -> TTC	102	SNP (transition)	None	100.0%	0.0
c	1,554	1,554	1	1		1,554	G -> C	GCG -> GCC	102	SNP (transversion)	None	100.0%	0.0
c	1,446	1,446	1	1		1,446	T -> C	ACT -> ACC	98	SNP (transition)	None	100.0%	4.9E-324
c	1,443	1,443	1	1		1,443	T -> C	TCT -> TCC	93	SNP (transition)	None	100.0%	6.3E-317
	1,377	1,377	1	1		1,377	T -> C	ATT -> ATC	96	SNP (transition)	None	100.0%	0.0
1	1,305	1,305	1	1		1,305	C -> T	GAC -> GAT	92	SNP (transition)	None	100.0%	2.5E-304
с	1,301	1,301	1	1	N -> T	1,301	A -> C	AAC -> ACC	89	SNP (transversion)	Substitution	100.0%	2.5E-303
r	1,257	1,257	1	1	K -> N	1,257	G -> T	AAG -> AAT	93	SNP (transversion)	Substitution	100.0%	0.0
r	1,245	1,245	1	1		1,245	C -> T	ATC -> ATT	103	SNP (transition)	None	100.0%	0.0
c	1.191	1.191	1	1		1.191	T -> C	AAT -> AAC	109	SNP (transition)	None	100.0%	0.0

Export table をクリックすると、テーブルをタブ区切り(.tsv)またはカンマ区切り(.csv)形式でエ クスポートすることができます。

次回はアノテーションの比較機能を使用して、より確度の高い SNP を検出する手法をご紹介する予 定です。

Geneious 製品概要・フリートライアルリクエストについては<u>こちら</u> 『Geneious Prime で猫も杓子もシークエンス解析』 過去の記事は<u>こちらでチェック!</u>

> TDB News 2. 2024 トミーデジタルバイオロジー株式会社 Phone 03-6240-0843 Fax 03-6240-0461