

# 次世代シーケンサーによる大規模転写産物解析の 自動アノテーションシステム:GIGANT

GIGANT: Automated Annotation System for Large-Scale Transcript Analyses Using Next Generation Sequencers

## ●研究の目的

●次世代シーケンサーの技術革新が目覚ましい昨今、シーケンサーから出力される大量配列情報をいかに処理して解析するかが問題となっている。解析には計算機による情報処理が不可欠だが、一般的な実験系研究室では、情報解析を適切に実施できる研究者・技術者が皆無、あるいは不足していることが主な要因である。我々は、この問題への解決策を提供するため、次世代シーケンサーからの転写産物の配列情報を入力として受け取ると、個々の配列に自動的にアノテーション情報を付与するシステムを構築した。

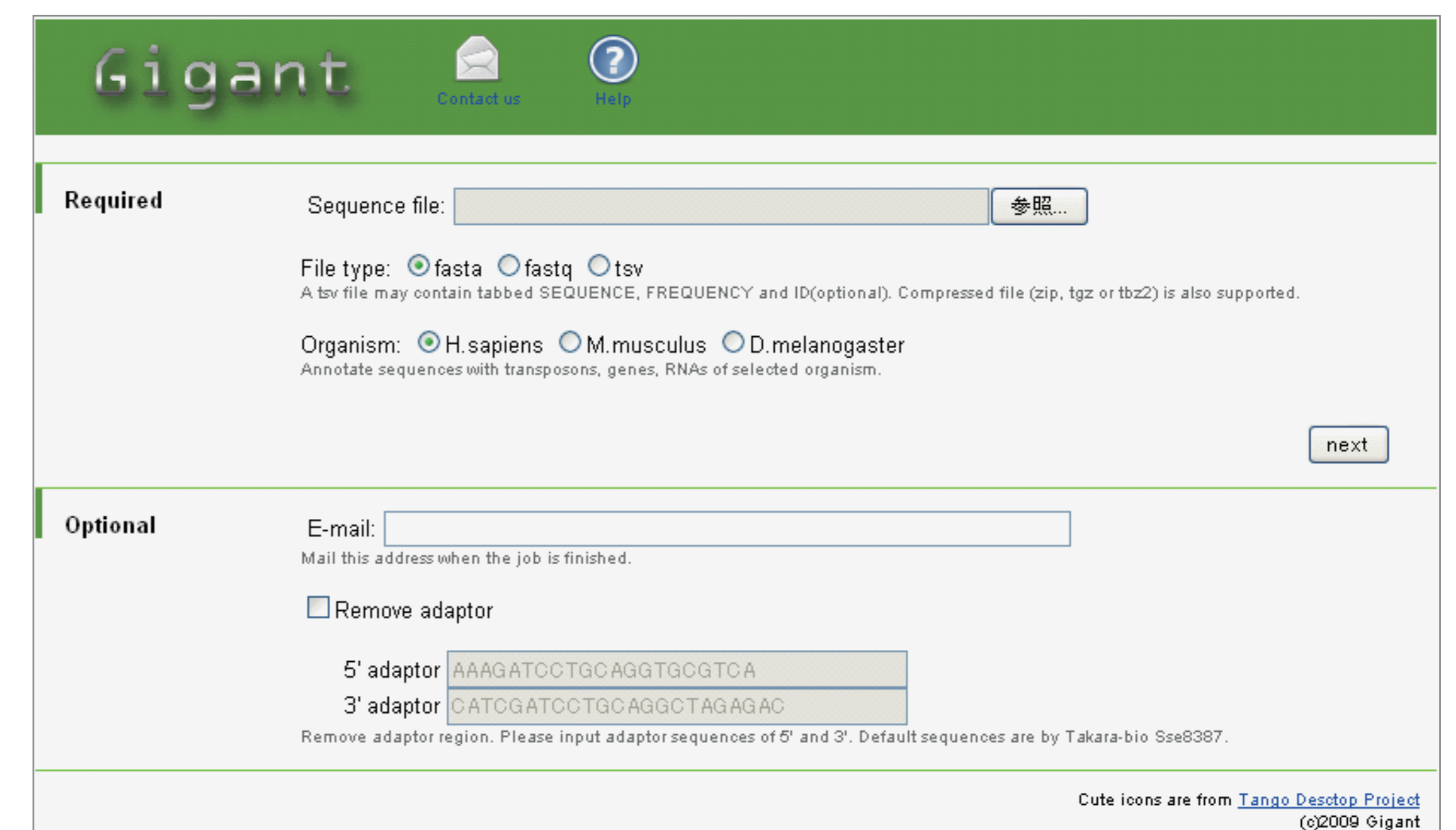
## ●研究の成果

●本システムは数百万配列規模の入力に対応している。シーケンサーが出力したFASTA形式ファイルあるいは、FASTQ形式のファイルをZIP等で圧縮された状態で本システムにアップロードする。入力配列に対して、トランスポゾンや繰返し配列を取り除いて、残った配列を分散処理によりゲノム配列にマッピング、転写の由来となった遺伝子を推定する。遺伝子の推定には、タンパク質遺伝子の情報だけでなく、機能性RNAデータベースで網羅的に収集した機能性RNAの情報も使用する。また、利用者がアノテーション処理の方法をいくつかの選択肢から選択することも可能で、利用者の目的に応じたアノテーションができるように工夫されている。現在対応済の次世代シーケンサーはイルミナ社のGAIIとRoche社のFLXである。また、対応しているゲノムは、ヒト、マウス、ラット、ショウジョウバエ。自動アノテーションシステムにはウェブサイト (<http://www.ncrna.org/>)にて公開予定である。

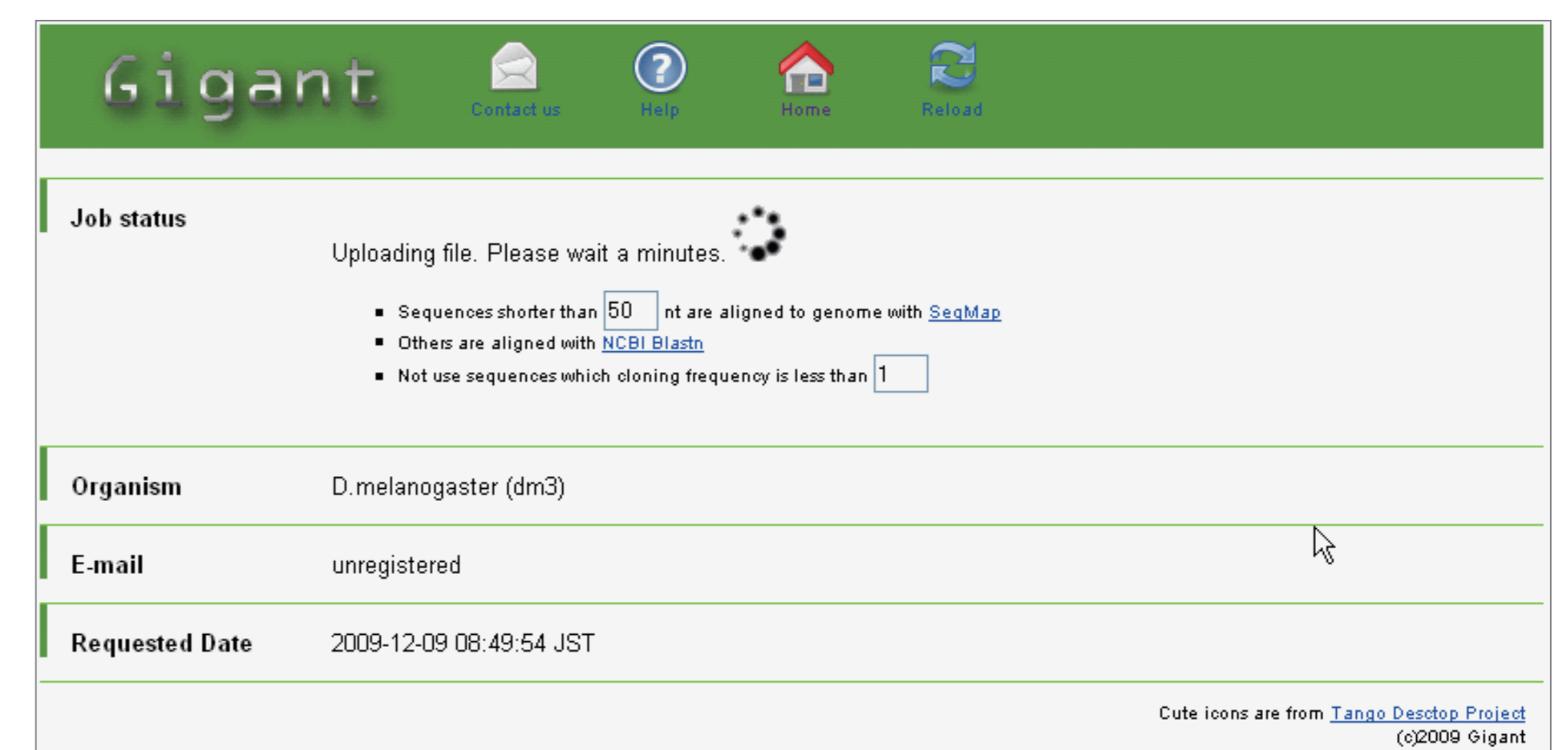
## ●今後の展開

●配列情報のアップロードにおいて、回線速度によっては十分な転送速度が得られない場合がある。一般的な圧縮プログラムよりも効率よく配列を圧縮することで短時間の配列情報の転送を実現するシステムも別途開発中である。さらに、アノテーション処理と、それ以降の処理について、利用者がワークフローを自由に設計できるよう、生命情報工学研究センターで開発中の、ワークフロー設計システムと連動させることを計画している。

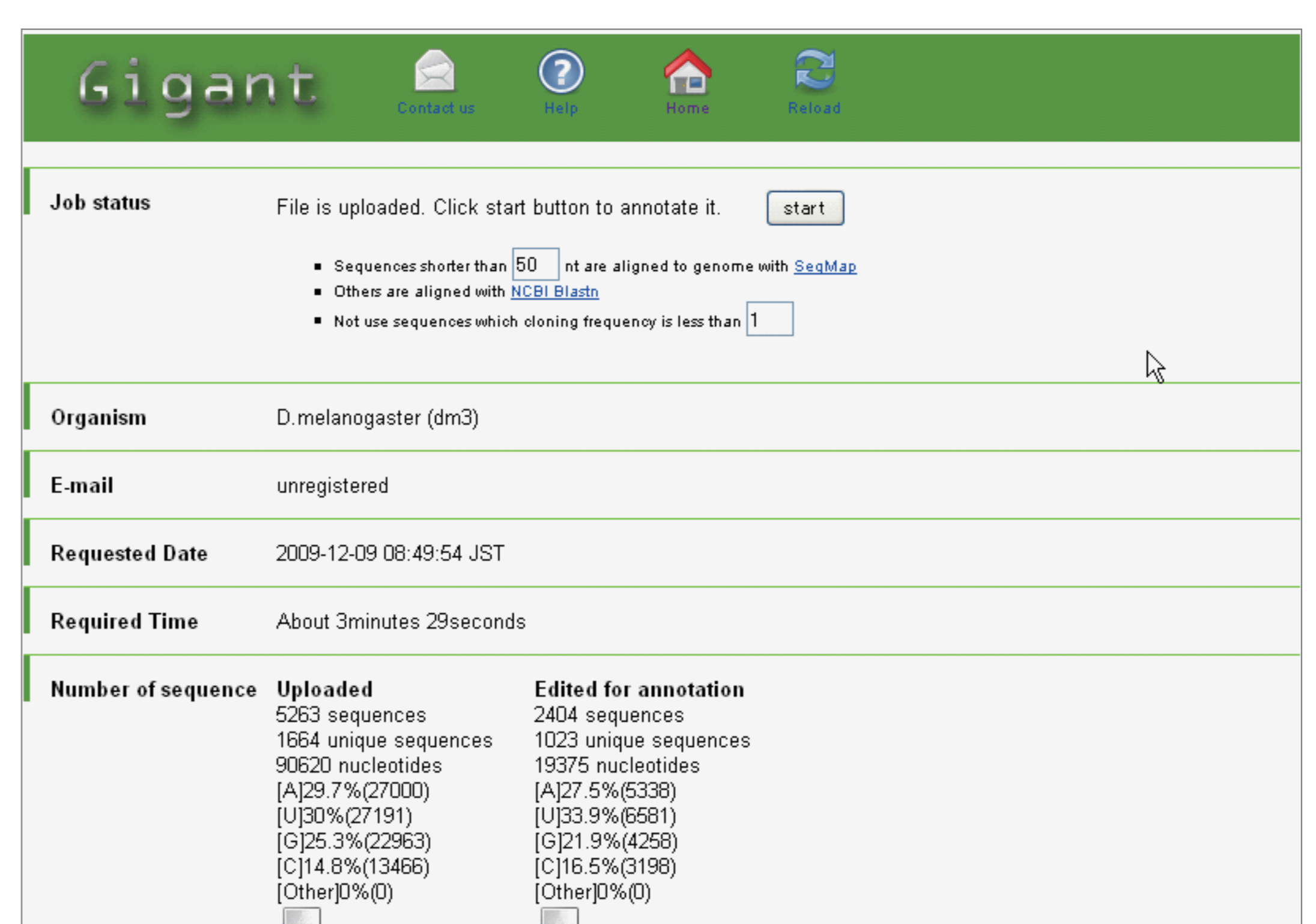
●本技術は、今後次世代シーケンサーが、臨床などより一般社会での利用に供される場合にも必要不可欠な技術である。



FAST形式やFASTQ形式の配列情報ファイルをZIPなどで圧縮された状態でアップロードする。



アップロードが完了すると、基本的な統計情報を表示して、これ以降のアノテーションに進むか確認する画面を表示する。



アノテーション結果はエクセルファイルで出力される。個々の配列が対応する遺伝子の情報が出力される。

id	original ID	C	D	E	F	G
1	unlike ID	reads	sequence	mapping quality	annotation	details
2	U0001	mim9_F_fly_smallRNA_FR487150	1	TGATGATGATGATGATAAA	repeat	(CAT)n, (ATG)n, CT-rich, (TAA)n, (TTA)n
3	U0002	mim9_F_fly_smallRNA_FR428923	1	ATGATGATGATGATGATAAA	repeat	(CAT)n, (ATG)n, CT-rich, (TAA)n, (TTA)n
4	U0003	mim9_F_fly_smallRNA_FR456449	1	GTGTTCTTGGACCTGGCCAT	repeat	(CAT)n, (ATG)n, (TAA)n, AT-rich, (TTA)n
5	U0004	mim9_F_fly_smallRNA_FR428919	1	ATGATGATGATGATGATAAA	RNA	(CA)n, (ATG)n, (TAA)n, (TTA)n
6	U0005	mim9_F_fly_smallRNA_FR492736	1	TGATGATGATGATGATAAA	RNA	(CA)n, (ATG)n, (TAA)n, (TTA)n
7	U0006	mim9_F_fly_smallRNA_FR446392	1	CTCTCCCCCCCCCCCCCG	repeat	(G)n, GC-rich, (C)n, G-rich, GA-rich
8	U0007	mim9_F_fly_smallRNA_FR426900	1	ATGATGATGATGATGATAAA	repeat	(CAT)n, (ATG)n, CT-rich, (CA)CAT)n, (ATG)G)n
9	U0008	mim9_F_fly_smallRNA_FR428919	1	CCCCCCCCCCCCCCCCCT	transposon	R1_DM_FW_DM_C_DM_D0C2_DM_ACCOR02_L
10	U0009	mim9_F_fly_smallRNA_FR425424	1	ATCATCATCATCATCAAAA	repeat	(CAT)n, (ATG)n, (TTG)n, (ATGG)G)n, (CACC)AT)n
11	U0010	mim9_F_fly_smallRNA_FR478824	1	TCATCATCATCATCAAAA	repeat	(CAT)n, (ATG)n, (TTG)n, (ATGG)G)n, (CACC)AT)n
12	U0011	mim9_F_fly_smallRNA_FR493645	1	TCCTGCTGCTGCGGATGGCC	repeat	(CAT)n, (ATG)n, (TTG)n, (ATGG)G)n, (CACC)AT)n
13	U0012	mim9_F_fly_smallRNA_FR428919	1	AAATGATGATGATGATAAA	transposon	BLOOD_I_BURDOCK_I_Gypsy4_L_DM297_I_TCI_DM
14	U0013	mim9_F_fly_smallRNA_FR473357	1	TAGTCAAT	RNA	IRNA_GCA(Cys)
15	U0014	mim9_F_fly_smallRNA_FR487144	1	TGATGATGATGATGATAAA	repeat	(CAT)n, (ATG)n, CT-rich, (CA)n, (TTA)n
16	U0015	mim9_F_fly_smallRNA_FR509733	1	TTTTTTCTTCTTTT	repeat	A-rich, T-rich, (T)n, (A)n, GA-rich
17	U0016	mim9_F_fly_smallRNA_FR428919	1	TTTATGATGATGATGATG	repeat	(CAT)n, (ATG)n, (TTG)n, (ATGG)G)n, (CACC)AT)n
18	U0017	mim9_F_fly_smallRNA_FR454740	1	GATGATGATGATGATGGA	repeat	(CAT)n, (ATG)n, (TTG)n, (ATGG)G)n, (CACC)AT)n, CT-rich
19	U0018	mim9_F_fly_smallRNA_FR454991	1	GATGATGATGATGATGGA	RNA	IRNA_TGC(Ala)
20	U0019	mim9_F_fly_smallRNA_FR427990	1	ATGATGATGATGATGATG	RNA	IRNA_TGC(Ala)
21	U0020	mim9_F_fly_smallRNA_FR489743	1	TGCTGGAGGAGGAGGAGG	repeat	(TC)n, CT-rich, (G)A)n, GA-rich, (CC)A)n
22	U0021	mim9_F_fly_smallRNA_FR465315	1	GTGGGGGGGGGGGGGGGG	repeat	(CC)A)n, (CC)A)n, (TGG)G)n, G-rich, (CTG)n
23	U0022	mim9_F_fly_smallRNA_FR424841	1	ATATGATGATGATGATGATG	repeat	(CAT)n, (ATG)n, CT-rich, (CA)n, (CACC)AT)n
24	U0023	mim9_F_fly_smallRNA_FR482070	1	TGATTTATTTT	transposon	ROO_I_Gypsy12A_LTR_ACCOR02_I_PROTOP_DM2
25	U0024	mim9_F_fly_smallRNA_FR428919	1	TCATCATCATCATCAAAA	repeat	(CAT)n, (ATG)n, (TTG)n, (ATGG)G)n, (CACC)AT)n
26	U0025	mim9_F_fly_smallRNA_FR459988	1	CGTGTGCTATGTGGCT	repeat	(CT)G)n, (CA)C)n
27	U0026	mim9_F_fly_smallRNA_FR454156	1	GATATAG	RNA	IRNA_GCA(Cys)

連絡先: RNA情報工学チーム 光山統泰 mituyama-toutai@aist.go.jp

<http://www.aist.go.jp/>