

Index

1. ファイルについて
2. プログラムの実行手順、実行コード例
3. インプット vcf の形式について
4. プログラムの出力について

普段どのような環境で分析されているのかわからないので、terminal と bcftools を使う場合のコードを記載しています。jupyter notebook でも実行可能です。

※ terminal で実行した場合、グラフ画像が生成された段階で実行が一時停止する可能性があります。その場合は画像(保存可能)を閉じると実行が再開されます。jupyter notebook 上で実行した場合はそのようなことは起きません。

※ 二つ目のグラフ画像の生成には十数秒かかる可能性があります。

1. ファイルについて

VCF_to_PhylotreeResult.py … レポート等を入力する関数が書かれているファイル。

Phylotree_df.pickle … VCF_to_PhylotreeResult.py の関数内で使われる、<https://www.phylotree.org> の情報をデータフレームオブジェクトとして保存したファイル

2. プログラム実行手順 (terminal 内でも実行可)

- ① Python 実行環境の構築、必要ライブラリのインストール (python, pip はインストール済と仮定)

(ア) pandas, numpy, matplotlib.pyplot, pickle の4つのパッケージのインストール

(例) terminal 内において `pip install pandas`

(イ) VCF_to_PhylotreeResult.py, Phylotree_df.pickle を含むファイルで入れ括りに移動

(例) terminal 内において `cd ディレクトリパス名`

(ウ) python をスタート

(例) terminal 内において `python`

- ② レポートを入力する関数のインポート

(例) terminal 内において `from VCF_to_PhylotreeResult import get_report`

- ③ レポートの取得

(例) `results_dataframe, filtered_dataframe, leaf_dataframe = get_report(“インプットファイルのディレクトリ名”)`

3. インプット vcf の形式について

CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO
-------	-----	----	-----	-----	------	--------	------

の順にデータが並んでいてヘッダー（コラム名を含め、# から始まる行）のない vcf ファイル

（例）bcftools の view -o -H メソッドで取得

4. プログラムの出力について（出力例）

Reference: <https://www.phylotree.org> <= 情報源

Credit:

<https://www.sciencedirect.com/science/article/abs/pii/S1875176815302432> <= 情報源の作成者による説明

-----Best Matching Haplogroup-----

グラフ画像 1 <= 変異の一致率が最も高い 50 種類のハプログループの棒グラフ

↓ グラフ画像 1 結果の説明文

You are most likely belonging to haplogroup ~

~% of your variant is matching haplogroup B4f

Match rate to ~ is 117.55675675675676 times the average match rate to other haplogroups

Links to examples of mtDNA samples in same haplogroup are listed below. It should include geographical/origin information of the sample.

リンク集 <= 同一のハプログループに属するサンプル情報へのリンク集

within the filtered bar graph, relationship between haplogroup is shown below:

矢印図 <= グラフ画像 1 に含まれているハプログループ同士の関係を示す矢印図 (phylotree.org に基づいた図)

-----Phylotree Paths-----

full relationship for non-zero matching rate haplogroups is shown below:

矢印図 <= 一致率が 0 でないハプログループ同士の関係を示す矢印図 (phylotree.org に基づいた図)

-----Best Matching Path-----

To see which leaf haplogroup (the leaf node groups in the phylotree) you belong to, variant-match-ratio for each path to the leaf haplogroup is computed

グラフ画像2 <= [phylotree.org](https://www.phylotree.org/) の木構造の葉ノードに分類されるハプログループに対する一致率（根ノードから各葉ノードまでの分岐に含まれる全ての変異に対して）の棒グラフ

↓ グラフ画像2 結果の説明文

You are most likely belonging to leaf haplogroup [~~~~](#)

[~~~~%](#) of your variant is matching with the path to [~~~~](#)

矢印図 <= 根ノードから一致率が最も高い葉ノードハプログループへの矢印図
([phylotree.org](https://www.phylotree.org/) に基づいた図)

Links to examples of mtDNA samples in same haplogroup are listed below. It should include geographical/origin information of the sample.

リンク集 <= 同一の葉ノードハプログループに属するサンプル情報へのリンク集