



Alignement de séquences et fiabilité

Projet tutoré
Master 1 - Bioinformatique

Etudiants:

CASTILLO Karina

CAUQUIL Delphine

CLEMENT Hannelore

FRANCOIS Philippe

Tuteurs:

RANWEZ Vincent

DENHE-GARCIA Alex

Index

I. Définition générale du problème

II. Vue générale du projet

Diagramme de Gantt

Diagramme de « cas d'utilisation »

Diagramme de « séquences »

Diagramme de relation des fonctions

III. Explication d'une fonction

IV. Conclusion

Définition générale du problème

>sequence1

ggg---gaag

>sequence2

ggg---gaag

>sequence3

gggcaggaag

>sequence4

ggg-----g

>sequence5

ggggaacagg

Définition générale du problème

Résolution :

- Comparaison des sites de l'alignement d'intérêt avec les sites d'alignements répliqués
- Calcul d'une matrice de fiabilité intermédiaire
- Calcul d'une matrice finale à partir de la moyenne des matrices intermédiaires

Définition du problème

Alignement référence

>sequence1 ggg---gaag	1	2	3	{3-4}	{3-4}	{3-4}	4	5	6	7
>sequence2 ggg---gaag	1	2	3	{3-4}	{3-4}	{3-4}	4	5	6	7
>sequence3 gggcaggaag	1	2	3	4	5	6	7	8	9	10
>sequence4 ggg-----g	1	2	3	{3-4}	{3-4}	{3-4}	{3-4}	{3-4}	{3-4}	4
>sequence5 ggggaacagg	1	2	3	4	5	6	7	8	9	10

Alignement réplikat

>sequence1 ggg---gaag	1	2	3	{3-4}	{3-4}	{3-4}	4	5	6	7
>sequence2 ggg---gaag	1	2	3	{3-4}	{3-4}	{3-4}	4	5	6	7
>sequence3 gggcaggaag	1	2	3	4	5	6	7	8	9	10
>sequence4 g-gg-----g	1	{1-2}	2	3	{3-4}	{3-4}	{3-4}	{3-4}	{3-4}	4
>sequence5 ggggaacagg	1	2	3	4	5	6	7	8	9	10

Recherche du site de référence dans le réplikat

Définition du problème

Alignement référence

>sequence1	1	2	3	{3-4}	{3-4}	{3-4}	4	5	6	7
ggg---gaag										
>sequence2	1	2	3	{3-4}	{3-4}	{3-4}	4	5	6	7
ggg---gaag										
>sequence3	1	2	3	4	5	6	7	8	9	10
gggcaggaag										
>sequence4	1	2	3	{3-4}	{3-4}	{3-4}	{3-4}	{3-4}	{3-4}	4
ggg-----g										
>sequence5	1	2	3	4	5	6	7	8	9	10
ggggaacagg										

Alignement répliat

>sequence1	1	2	3	{3-4}	{3-4}	{3-4}	4	5	6	7
ggg---gaag										
>sequence2	1	2	3	{3-4}	{3-4}	{3-4}	4	5	6	7
ggg---gaag										
>sequence3	1	2	3	4	5	6	7	8	9	10
gggcaggaag										
>sequence4	1	{1-2}	2	3	{3-4}	{3-4}	{3-4}	{3-4}	{3-4}	4
g-gg-----g										
>sequence5	1	2	3	4	5	6	7	8	9	10
ggggaacagg										

Recherche du site de référence dans le répliat

Définition du problème

Alignement référence

>sequence1										
ggg---gaag	1	2	3	{3-4}	{3-4}	{3-4}	4	5	6	7
>sequence2										
ggg---gaag	1	2	3	{3-4}	{3-4}	{3-4}	4	5	6	7
>sequence3										
gggcaggaag	1	2	3	4	5	6	7	8	9	10
>sequence4										
ggg-----g	1	2	3	{3-4}	{3-4}	{3-4}	{3-4}	{3-4}	{3-4}	4
>sequence5										
ggggaacagg	1	2	3	4	5	6	7	8	9	10

Alignement réplikat

>sequence1										
ggg---gaag	1	2	3	{3-4}	{3-4}	{3-4}	4	5	6	7
>sequence2										
ggg---gaag	1	2	3	{3-4}	{3-4}	{3-4}	4	5	6	7
>sequence3										
gggcaggaag	1	2	3	4	5	6	7	8	9	10
>sequence4										
g-gg-----g	1	{1-2}	2	3	{3-4}	{3-4}	{3-4}	{3-4}	{3-4}	4
>sequence5										
ggggaacagg	1	2	3	4	5	6	7	8	9	10

Comparaison des colonnes

Définition générale du problème

Matrice de fiabilité (Référence vs réplikat)

1	0.8	0.8	-	-	-	1	1	1	1
1	0.8	0.8	-	-	-	1	1	1	1
1	0.8	0.8	0.8	1	1	1	1	1	1
1	0.2	0.2	-	-	-	-	-	-	1
1	0.8	0.8	0.8	1	1	1	1	1	1

Matrice intermédiaire

Vue générale du projet

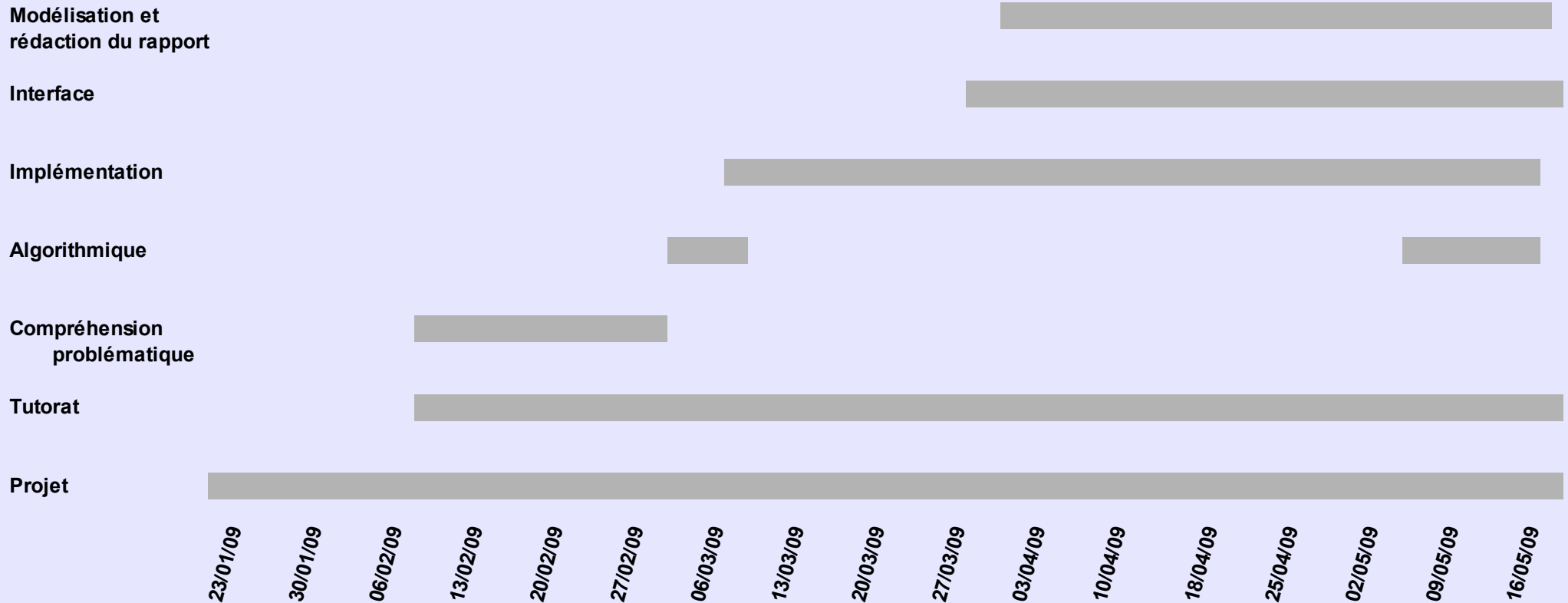


Diagramme de Gantt

Vue générale du projet

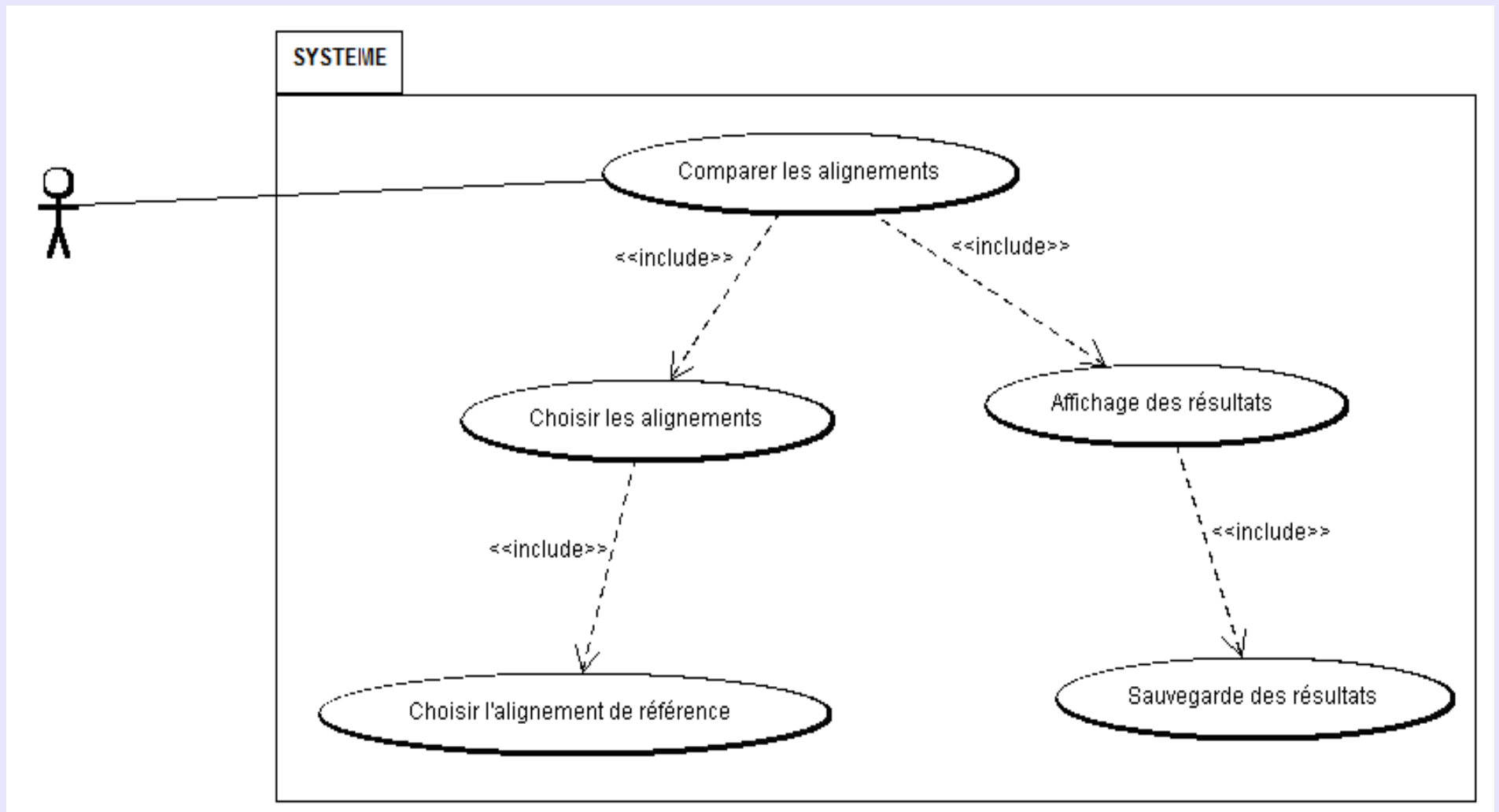


Diagramme « cas d'utilisation »

Vue générale du projet

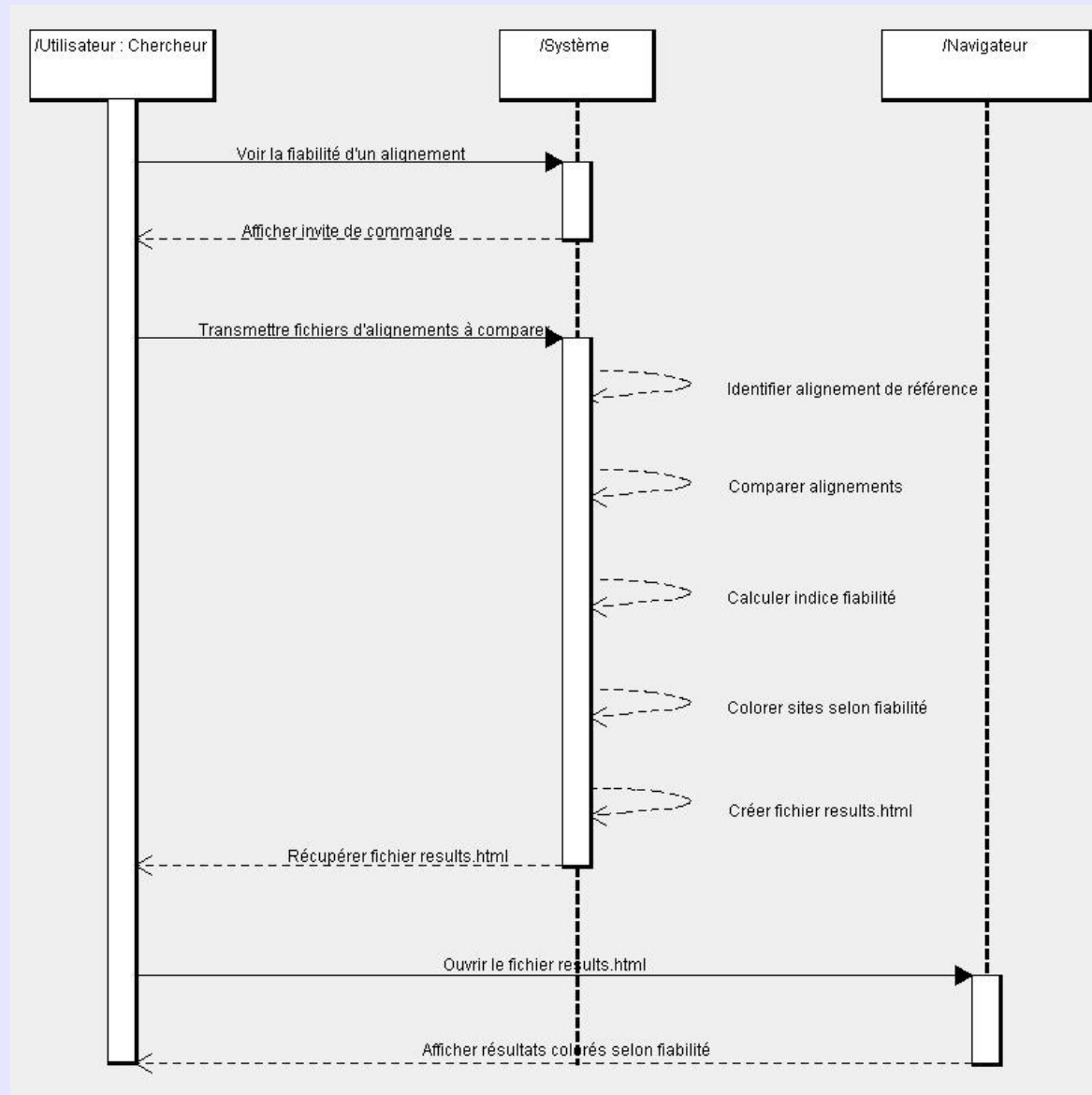


Diagramme de « séquences »

Vue générale du projet

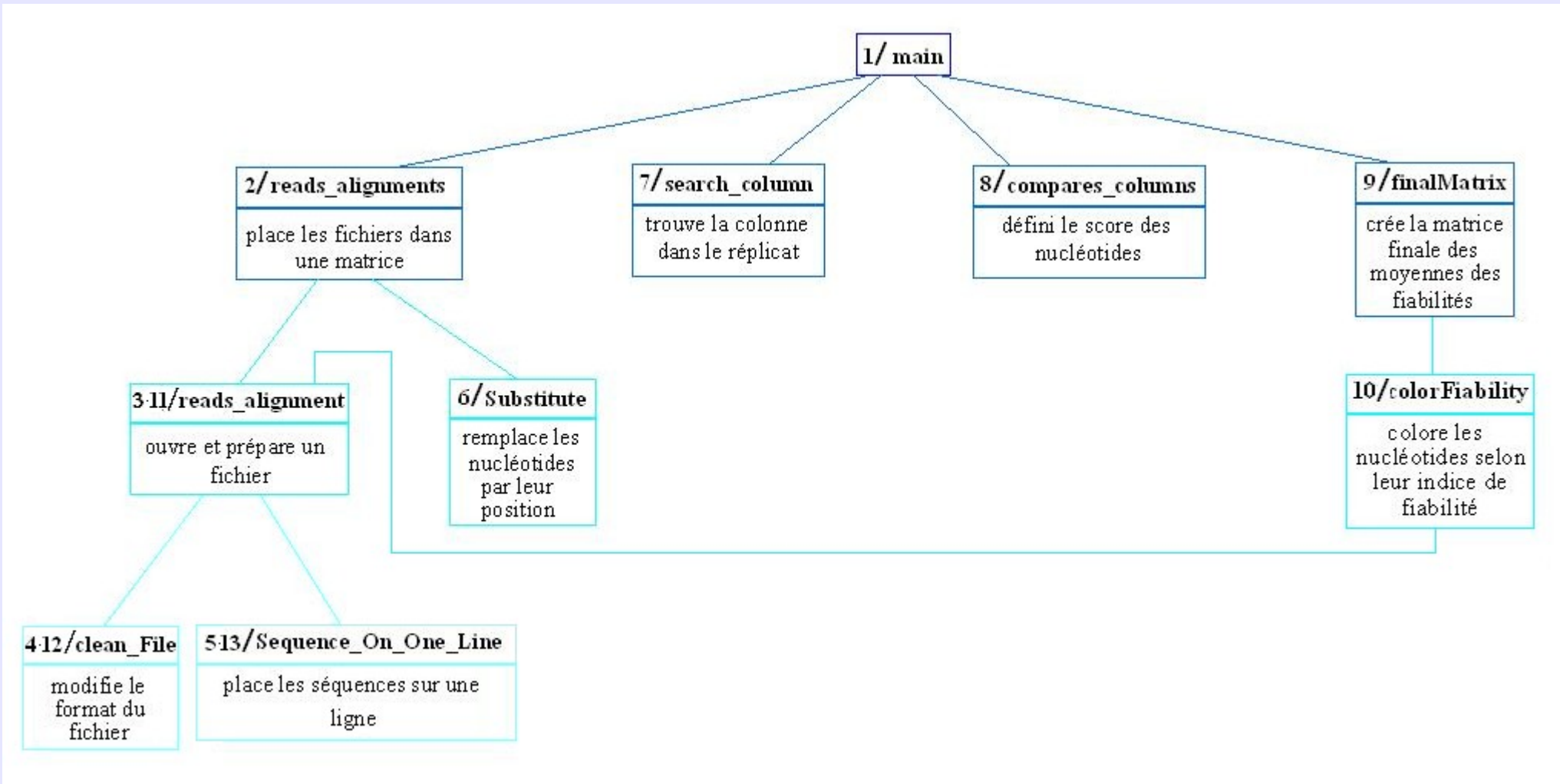


Diagramme de relation des fonctions

Code de la fonction Substitute

```
#substitute the nucleotid with its position in the sequence
sub Substitute {
  my %matrix;
  my @enter_files_fasta = @_;
  my $seq_name=0;
  for my $line (@enter_files_fasta) {
    if ($line =~ /^>(.)+/) {
      $seq_name = $1;
    }
    elsif ($line =~ /^(.+)/i) {
      my $count=0;
      my @columns;
      for my $column (split //, $1) {
        if ($column =~ /\w/i) {
          $count++;
          ##@columns contains the positions
          push @columns, "$seq_name$count";
        }
        else {
          push @columns, "$seq_name\${count-@{[$count+1]}\}";
        }
      }
      @{$matrix{$seq_name}} = @columns;
    }
  }
  return %matrix;
}
```

```
>mammifere
gg----gaag
>humain
ggg----aag
>jambon
gggcaggaag
>serpent
ggg-----g
>souris
ggggaacagg
```

```
humain :          jambon :
humain1          jambon1
humain2          jambon2
humain3          jambon3
humain<3-4>     jambon4
humain<3-4>     jambon5
humain<3-4>     jambon6
humain<3-4>     jambon7
humain4          jambon8
humain5          jambon9
humain6          jambon10

serpent :        mammifere :
serpent1         mammifere1
serpent2         mammifere2
serpent3         mammifere<2-3>
serpent<3-4>    mammifere<2-3>
serpent<3-4>    mammifere<2-3>
serpent<3-4>    mammifere<2-3>
serpent<3-4>    mammifere3
serpent<3-4>    mammifere4
serpent<3-4>    mammifere5
serpent4         mammifere6

souris :
souris1
souris2
souris3
souris4
souris5
souris6
souris7
souris8
souris9
souris10
```

```
humain1          souris1          serpent1
humain2          souris2          serpent2
humain3          souris3          serpent3
humain<3-4>     souris4          serpent4
humain<3-4>     souris5          serpent<4-5>
humain<3-4>     souris6          serpent<4-5>
humain<3-4>     souris7          serpent<4-5>
humain4          souris8          serpent<4-5>
humain5          souris9          serpent<4-5>
humain6          souris10         serpent<4-5>
mammifere1      jambon1
mammifere2      jambon2
mammifere3      jambon3
mammifere<3-4>  jambon4
mammifere<3-4>  jambon5
mammifere<3-4>  jambon6
mammifere<3-4>  jambon7
mammifere4      jambon8
mammifere5      jambon9
mammifere6      jambon10
```

Conclusion

Alignment fiability

AAvianOuzbekistan2008H5N1 :

```
-----TGGAATATGGTAACTGCAACACCCAAGTGTCAAACCTCCAATAGGGGGCGAT
AAACTCTAGTATGCCATTCCACAACATCCACCCTCTCACCATTGGGGAATGCCCCA
AATATGTGAAATCAAACAGATTAGTCCCTTGCGACTGGGCTCAGAAATAGCCCTCAA
GGGGAGAGAAGAAGAAAAA-----GAGAGGACTATTTGGAGCC-----AT
AGCAGGTTTTATAGAG-----GGAGGATGGCAGG-----GCATGGTAGAC-----
-----GGTTGG-----
```

AAvianTurkey2007H5N1 :

```
-----CCATTCCACA
ACATACACCCTCTCACCATCGGGGAATGCCCAAATATGTGAAATCAAACAGATTAA
GTCCTTGCGACTGGGCTCAGAAATAGCCCTCAAGAAGAGAGAGAAGAAGAAAAA----
---GAGAGGACTATTTGGAGCT-----ATAGCAGGTTTTATAGAG-----GG
AGGATGGCAGG-----GAATGGTAGAT-----GGTTGGTATGGGTACCACCAT
AGCAATGAGCAGGGG-----
```

AHumanThailande2009H1N1 :

```
CAGITCAGATACACCAGTCCACGATTGCAATACAACTTGTTCAGACACCCAAAGGGTG
CTATAAACACCAGCCTCCCAITTTCAGAAATATACATCCGATCACAAITGGAAAAATGTC
CAAAATATGTAAAAAGCACAAAATTGAGACTGGCCACAGGATTGAGGAAATGTCCCG
TCTATT-----CAATC-----TAGAGGCCTATTTGGGGCC-----ATTGGC
CGGTTTCATTGAA-----GGGGGGTGGACAG-----GGATGGTAGAT-----G
GATGGTACGGGITATCACCATCAAAATGAGCAGGGGGTCAGGAT
```

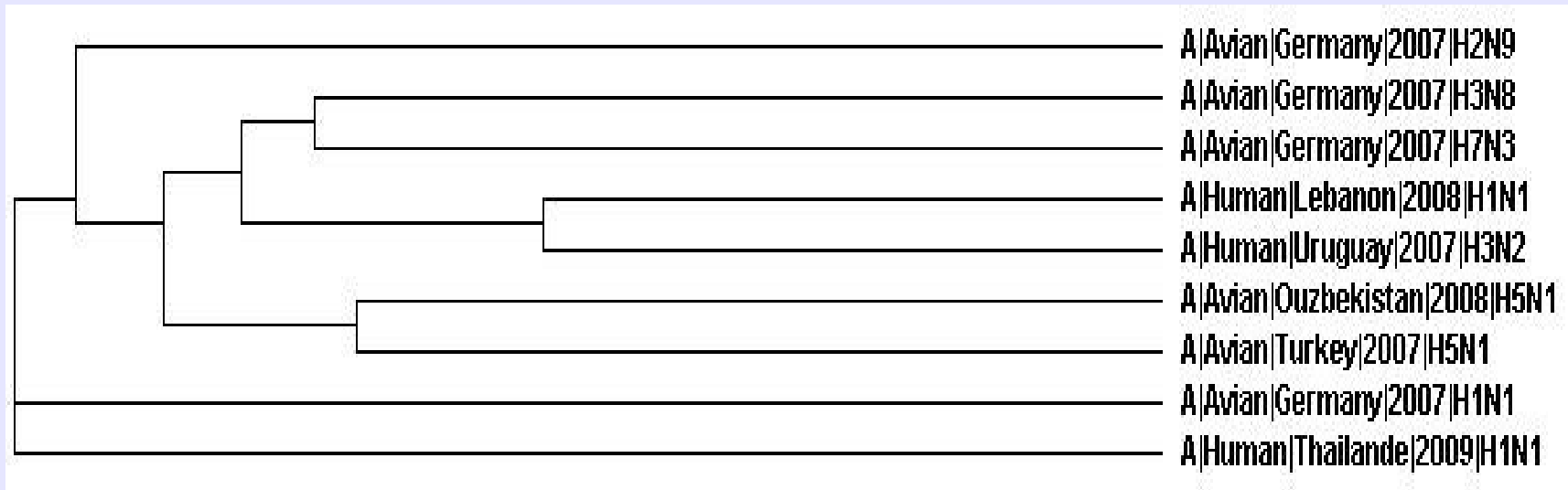
AHumanUruguay2007H3N2 :

```
-----TAATAACAGTTTCTTTAGTAGATTGAATTGGTTGACCCACTTAAAA
TTCAAATACCCAGCATTGAACGTGACTATGCCAAACAATGAAAAATTTGACAAAT-T
GTACATTTGGGGGGTTCCACCACCCGGGTACGGACAATGACCAAATCTTCCTGTA----
---TGCTC-----AAG-----CATCAGGAAAGA-----ATCAGAGTCTCTAC
CAAA-----AGAAAGCCAAACAAACTGTAATCCCGAATATCGGATCTAGACCCAGAG
TAAGGAATATCCCAGCAGAATAAGCATCTATTGG-----
```

Pourcentage de fiabilité :

tranche des 0% | tranche des 10% | tranche des 20% | tranche des 30% | tranche des 40% | tranche des 50% | tranche des 60% | tranche des 70% | tranche des 80% | tranche des 90% | tranche des 100% | gap.

Conclusion



- Création d'une interface homme-machine pour une utilisation plus simple

**Nous vous remercions de votre
attention.**




Fiabilité d'alignement de séquence ADN

ETAPE 1 / 2

Définition du nombre d'alignements à comparer :

envoyer



F.A.S

Fiabilité d'alignement de séquence ADN

ETAPE 2 / 2


Vos alignements doivent être en format FASTA

Votre fichier pour l'alignement de référence

Votre fichier pour l'alignement 1

Votre fichier pour l'alignement 2

Votre fichier pour l'alignement 3



Alignment fiability

A/HumanLebanon2008H1N1 :

```
.....TTCCCCAAA  
GAAAGCTCATGGCCCAACCCACACCGTAACCGGAGTGTGTCAGCATCATGCTCCCATAA  
TGGGGAAAGCAATTTTTACAGAAATTTGCTATGGCTGACGGGGAA.....GAATG  
GTTTGTACCA
```

AAvianGermany2007H1N1 :

```
.....CCTGAGAAAGATTCC  
TCCATT.....CAATC.....TAGAGGGCTCTTGGGAGCA.....ATTG  
CTGGTTTATTGAA.....GGGGGTGGACAG.....GAATGATAGAT.....  
GG
```

AAvianGermany2007H2N9 :

```
.....TTGAGAAATGTCCCT  
CAGATT.....GAATC.....AAGAGGACTGTTTGGAGCA.....ATTG  
CTGGTTTATTGAA.....GGAGGATGGCAAG.....GAATGGTTGAT.....  
GGTT
```

AAvianGermany2007H3N8 :

```
.....CAGGAATGGGGAATGTAC  
CAGAGAAA.....GAAAC.....GAGAGGCCATTGGGAGCA.....AT  
AGGAGGCTTATTGAA.....AATGGATGGGAGG.....GAATGATAGAT.....  
GGTT
```

AAvianGermany2007H7N3 :

```
.....CAGGAATGAAGAAATGTC  
CCGAAATC.....CCAAAGGG.....AAGAGGCCATTGGTGGT.....A  
TAGGGGTTTATTGAA.....A.....AT
```

AAvianOuzbekistan2008H5N1 :

```
.....TGGAAATATGGTAACTGCAACACCAAGTGTCAAACCTCCAATAGGGGGGAT  
AAACTCTAGTATGCCATTCCACAACATCCACCCTCTCACCATTGGGGAAATGCCCCA  
AATATGTGAAATCAAACAGATTAGTCTTGGCGACTGGGCTGAGAAATAGCCCTCAA  
GGGGAGAGAAGAAAGAAAAA.....GAGAGGACTATTGGAGCC.....AT  
AGCAATGAGCAAG.....GGAGGATGGCAGG.....GCATGGTAGAC  
GGTTGG
```

AAvianTurkey2007H5N1 :

```
.....GCATTCCACA  
ACATACACCCCTCTCACCATCGGGGAATGCCCAAAATATGTGAAATCAAACAGATTA  
GTCCTTGGGACTGGGCTCAGAAATAGCCCTCAAGAAGAGAGAAAGAAAAA.....  
GAGAGGACTATTGGAGCT.....ATTGGAGGTTTTATTAGAG.....GG  
AGGATGGCAGG.....GAATGGTAGAT.....GGTTGGTATGGGTACCCACCAT  
AGCAATGAGCAGGGG
```

AHumanThailande2009H1N1 :

```
.....GCATTCCACA  
CAGTTCAGATACACCAGTCCACGATTGCAATACAAGTTGTGAGACACCCAAGGGTG  
CTATAAACACACAGCCCTCCCATTTCAGAATATACATCCGATCACAATTGGAAAAATGT  
CAAAATATGTAAAAAGCAGAAAAATGAGACTGGCCACAGGATGAGGAAATGTCCCG  
TCTATT.....CAATC.....TAGAGGCCATTGGGGCC.....ATTGG  
GGTTTATTGAA.....GGGGGTGGACAG.....GCATGGTAGAT.....G  
GATGGTACGGTTATCAGCATCAAAAATGAGCAGGGGTGAGGAT
```

AHumanUruguay2007H3N2 :

```
.....TAATAACAGTTTCTTTAGTAGATTGAATTGGTTGACCCACTTAAAA  
TTCAAATACCCAGCATTGAACGTGACTATGCCAAACAATGAAAAATTTGACAAAT-T  
GTACATTTGGGGGTTCAACCACCGGTACGGACAATGACCAAATCTTCCTGTA...  
TGCTC.....AAG.....GATCAGGAAGA.....ATCAAGTCTTTAC  
GAA.....AGAGGCCAACAAAATGTAATCCCGAATATCGGATCTAGACCCAGAG  
TAAGGAATATCCCCAGCAGAAATAAGCATCTATTGG
```

Pourcentage de fiabilité :

tranche des 0% | tranche des 10% | tranche des 20% | tranche des 30% | tranche des 40% | tranche des 50% | tranche des 60% | tranche des 70% | tranche des 80% | tranche des 90% | tranche des 100% | gap.

Alignment fiability

ceci est contenu dans le tableau FileName :

Alik1.txt Alik2.txt Alik3.txt Alik4.txt

```
$VAR1 = [ 'Alik1.txt', 'Alik2.txt', 'Alik3.txt', 'Alik4.txt' ]; Alik1.txtAlik2.txtAlik3.txtAlik4.txtAlik1.txt$VAR1 = {};
```

Pourcentage de fiabilité :

tranche des 0% | tranche des 10% | tranche des 20% | tranche des 30% | tranche des 40% | tranche des 50% | tranche des 60% | tranche des 70% | tranche des 80% | tranche des 90% | tranche des 100% | gap.