BIOINFORMATICS TOOLS FOR EVALUATING CELIAC DISEASE AND ITS RISKS

Alexandra Čičmancová
University of Tartu
Bioinformatics seminar
25.11.2016
I HAVE NO IDEA WHAT I'M DOING
WE WILL GO THROUGH

1. Celiac disease
2. Novel protein risk assignment tool
3. Fasta bioinformatics tool
4. Homework
WHAT IS CD?

(GLUTEN-SENSITIVE ENTEROPATHY, CELIAC SPRUE)

• genetically linked inflammatory immune disease

• affects ~ 1.5% of the population

• symptoms after the consumption of food containing proteins from wheat, barley, rye and possibly oats and other grass family grains closely related to wheat

• symptoms (usually associated with the digestive tract): chronic diarrhea, abdominal pain, cramping, bloating or irritable bowel syndrome, general nutritional deficiency, failure to thrive, mouth ulcers, and fatigue

• leads to flattening of the villi in the small intestine, erosion of the mucosal epithelium
CD II.

- somewhat genetically determined
- 90 – 95 % - HLA-DQ2 encoded by alleles DQA1 * 05 and DQB1 * 02
- 5-10 % - HLA-DQ8 encoded by alleles DQA1 * 03 and DRB1 * 0302nd
BIOINFORMATICS

• should provide efficient evaluation to demonstrate:
  • the SAFETY of >98+% of wheat proteins that would not cause CD
  • RISK ~ 100% identification of those that would cause CD
NOVEL PROTEIN RISK ASSESSMENT TOOL

- tool to identify Exact Peptide matches between the amino acid sequence of a query protein and the 1,016 naturally occurring, mutated or determined peptides from wheat and wheat relatives

- subfamily Pooideae of the grass family (Poaceae) may pose a risk if they contain celiac active peptides

- [http://www.allergenonline.org/celiacbrowse.shtml](http://www.allergenonline.org/celiacbrowse.shtml)
FASTA

• Developed by W.R. Pearson (Virginia)
• To compare the query protein against 68 celiac including proteins
• adds level of safety
• identity matches <45%
• E score > 1*10^{-15}

http://www.allergenonline.org/celiacfasta.shtml

unlikely to present a risk of inducing celiac disease
QUESTIONS?
HOMEWORK

• use the database
  (http://www.allergenonline.org/databas ebrowse.shtml) to find protein sequence in FASTA format

• use FASTA tool (use Full Fasta)
  (http://www.allergenonline.org/celiacfa sta.shtml) to find out whether the protein is unlikely to present a risk of inducing celiac disease

• Describe the result in few sentences

<table>
<thead>
<tr>
<th>Name</th>
<th>gi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anna Ufliand</td>
<td>56000996</td>
</tr>
<tr>
<td>Slue Reisberg</td>
<td>21755</td>
</tr>
<tr>
<td>Liis Kolberg</td>
<td>6073860</td>
</tr>
<tr>
<td>Ahto Salumets</td>
<td>17907758</td>
</tr>
<tr>
<td>Marili Palover</td>
<td>114091</td>
</tr>
<tr>
<td>Kreete Lüll</td>
<td>1699225</td>
</tr>
<tr>
<td>Timo Tõnis Sikka</td>
<td>84029333</td>
</tr>
<tr>
<td>Andrii Rozumnyi</td>
<td>585290</td>
</tr>
<tr>
<td>Gagandeep Singh</td>
<td>171853012</td>
</tr>
<tr>
<td>Ants-Oskar Mäesalu</td>
<td>736319</td>
</tr>
<tr>
<td>Sander Tars</td>
<td>89892725</td>
</tr>
<tr>
<td>Jakub Dobrik</td>
<td>75140047</td>
</tr>
<tr>
<td>Yuri Toma</td>
<td>1313966</td>
</tr>
<tr>
<td>Artem Bachynskyi</td>
<td>548948848</td>
</tr>
</tbody>
</table>
Please send your homework to
sasa.cicmancova@gmail.com
by 02.12. 10:00