Virulence Factor Activity Relationship Group

Update for the NDWAC CCL Work Group

Plenary Meeting

December 16-17, 2002

VFAR Activity Group Participants

- Jeff Griffiths
- Graciela Ramírez Toro
- O. Colin Stine
- Others
  - EPA: Paul Berger, Jim Sinclair, Nena Nwachuku, Jafrul Hassan
  - Cadmus: Jo Anne Shatkin, Nelson Moyer, Lalitha Subramanian, Susan Bjork, Trudy Wassenaar

VFAR Activity Group: Initial Activities

OBJECTIVE

- Determine the availability of data for developing the VFAR concept for future CCLs.

TASK

- Conduct keyword searches on GenBank and the Comprehensive Microbial Resource (CMR -TIGR (The Institute of Genomic Research) databases as a first step to identify data availability.

  - Initial evaluation of available information in **GenBank** and **CMR** to determine whether these could be used for VFAR

VFAR Activity Group: Initial Activities

- **Discussed lists of organisms for searches;**
  - Which are the pathogens of concern?
    - Waterborne
- **Create glossaries with keywords for searches;**
  - Which keywords to use?
    - Words most likely to be used in entering data to the databases which describe *pathogenicity* and *virulence*.
      - The group suggested 17 keywords

VFAR Activity Group: Initial Activities (continuation)

Keywords

- pathogen
- pathogenicity
- resistance
- resistant
- toxin
- virulence

Pathogens

- *Campylobacter jejuni*
- coxsackievirus
- *Cryptosporidium parvum*
- hepatitis
- *Helicobacter*
- *Salmonela*

VFAR Activity Group: Initial Activities (continuation)

- Grid (narrowing search results)

	diarrhea	infection or infections or infectious	resistance	resistant	toxic	toxin or toxins	virulence	Total # of hits using keywords, limited by organism	Total hits without keywords, limited by organism
Acanthamoeba	0	39	0	0	0	0	1	40	669
Aeromonas hydrophilia	0	2	14	1	0	12	4	33	243
Bacillus anthracis	0	5	3	0	0	9	19	36	278
Balantidium coli	0	0	0	0	0	0	0	0	2
Blastocystis hominis	0	0	0	0	0	0	0	0	26
Campylobacter jejuni	0	39	27	1	0	4	11	82	595
coxsackievirus	0	3	0	0	0	0	0	3	22
Cryptosporidium parvum	0	205	12	0	0	1	0	218	8375
Cyanobacteria	0	2	36	12	132	12	12	206	6837
enterovirus	0	1412	0	2	0	0	1	1415	4220
Escherichia coli	17	541	408	89	4	438	355	1852	11503
Giardia	0	67	4	11	0	1	0	83	2963
Helicobacter pylori	0	162	129	47	0	0	160	498	3813
hepatitis A	0	251	0	1	0	0	0	252	0
hepatitis E	0	153	0	0	0	0	0	153	0
Legionella pneumophila	0	43	8	0	0	1	20	72	422
Norwalk virus and NLVs	8	39	0	0	0	0	0	47	946
polio virus	0	0	0	0	0	0	0	0	1024

VFAR Activity Group: Problems encountered

- Keyword searches are not always productive because there is no standard terminology
  - No systematic way of entering data
- The grid reduces the number of hits, but is still limited by selected pathogen and keywords
  - Unknown or unidentified waterborne pathogens are excluded
- GenBank can produce redundant results

VFAR Activity Group: Problems encountered (continuation)

- Sequence searches confounded by fragments vs. whole genes
  - CMR (TIGR) limited to the entire genome
  - Database searches with whole genomes may be more productive, if enough organisms are available.
- Lots of potential but these systemic problems, and possibly others, need to be addressed

VFAR Activity Group: Next Steps

- Conduct keyword searches to identify virulence, colonization, persistence genes from existing genomes.

- Conduct sequence searches for homology to identify all virulence, colonization, persistence genes from existing databases.

- Committee to review Wassenaar proposal for comparative genetics approach to identify intersection of waterborne and pathogenic organisms using bioinformatics on a supercomputer.