

Machine learning-based typing of *Clostridium botulinum* by FT-IR spectroscopy: preliminary results

Drigo I¹, Guolo A¹, Rizzardi A.¹, Cordovana M.², Palazzolo L.¹, Garbuio M.¹, Cordioli B.¹, Zandonà L.¹, Bano L.¹

¹ Istituto Zooprofilattico sperimentale delle Venezie, SCT2- Laboratory of Treviso; ² Bruker Daltonics GmbH & Co, Bremen Germany

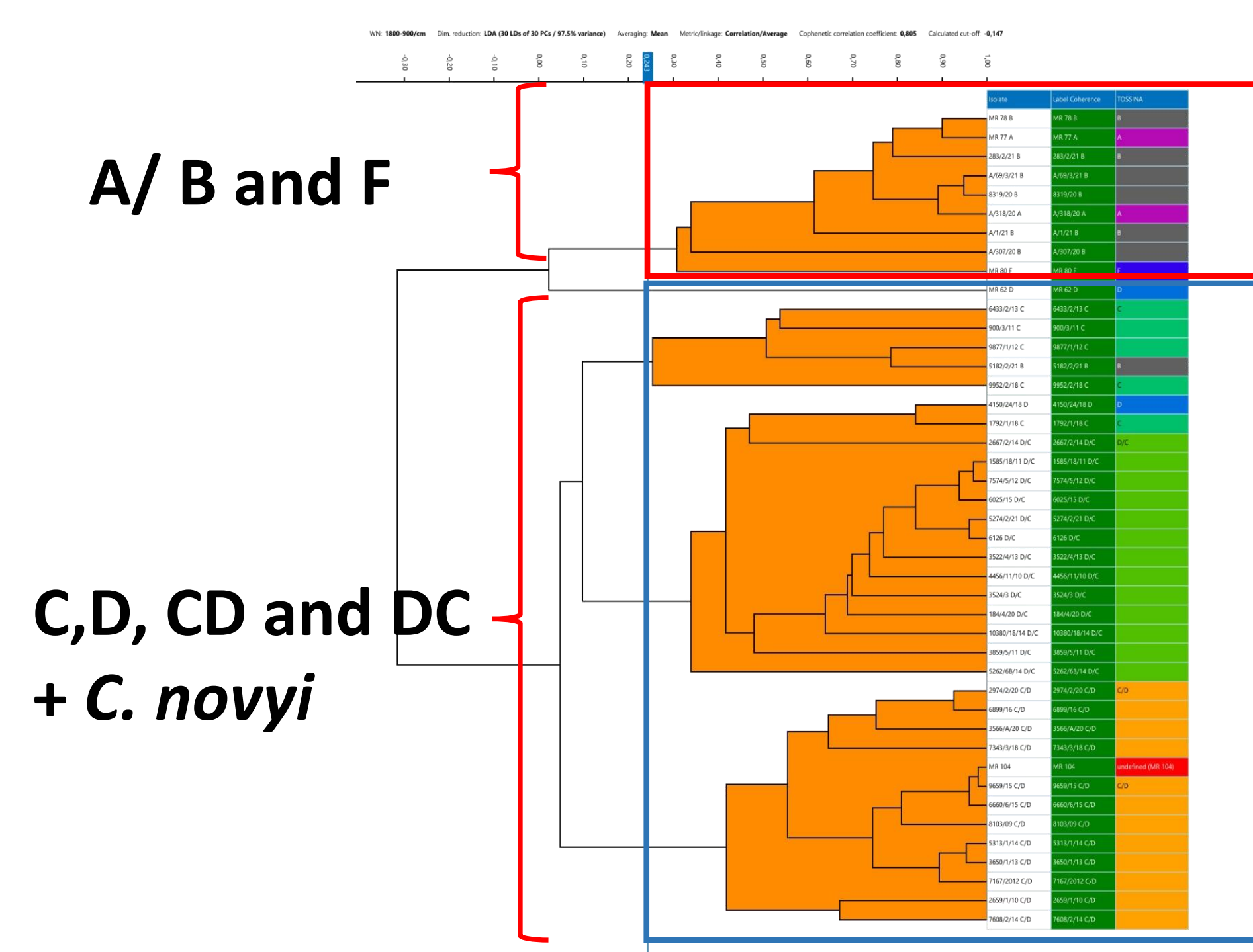
AIM: to investigate the use of FT-IRS for *C. botulinum* differentiation with respect to the botulinum neurotoxin encoding-gene type.

STRAINS: 48 ± 2 h cultures of 41 BoNT producing strains (BPS) at 37 ± 2 ° C

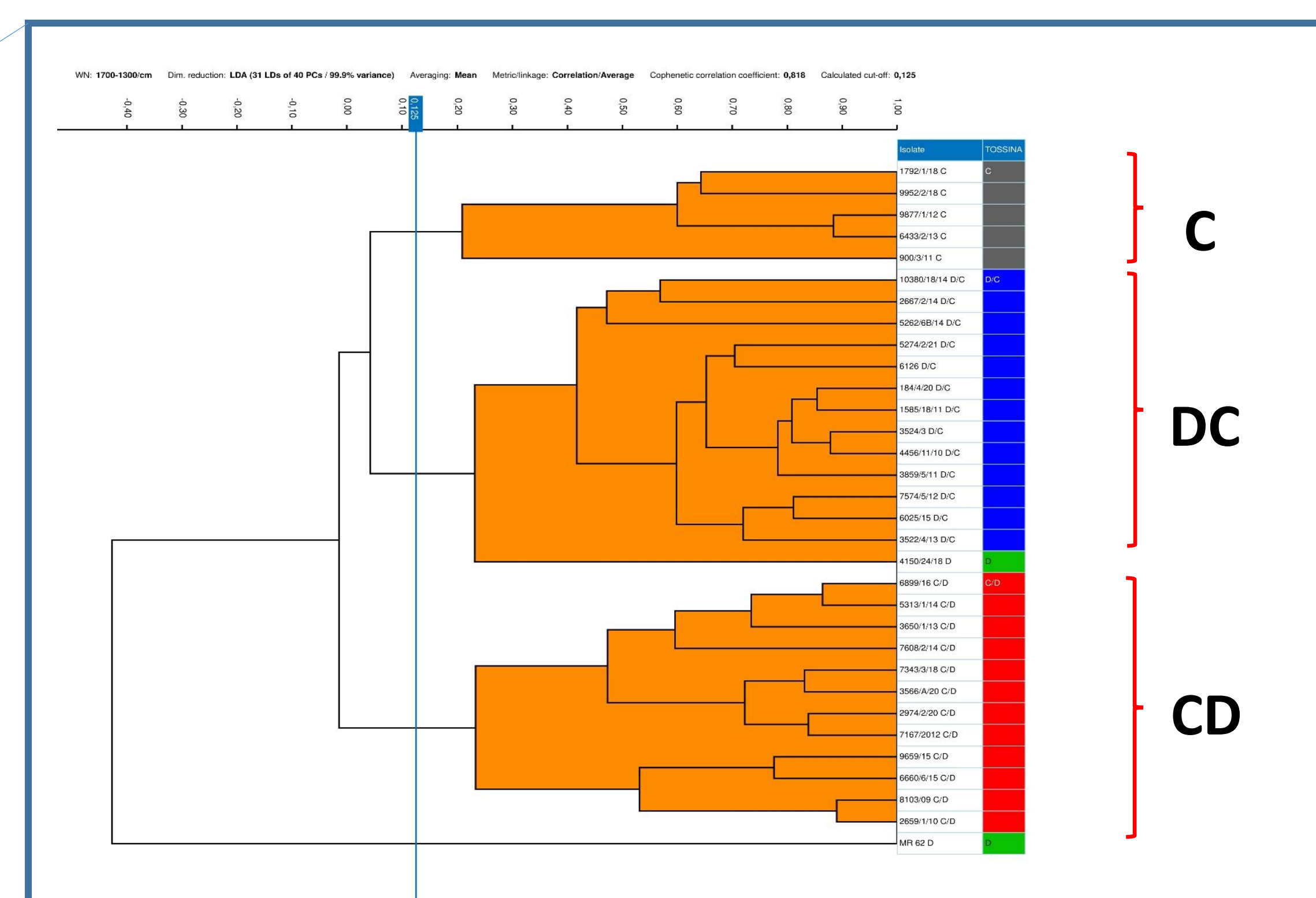
- 5 reference strains (ATCC 19397 type A, ATCC 27765 type B, NCTC 10281 type F, NCTC 8265 type D/C, *C. novyi* ATCC 25758)
- 32 animal field strains: 5 type C, 1 D, 12 type C/D and 13 D/C
- 6 field strains isolated from human outbreaks and foods: 1 type A (vegetables) and 6 type B (3 human feces, 2 cooked vegetables, 1 ready to eat legume soup)

BACTERIAL HANDLING: EtOH/water 1:1 cell component extraction suspension; sample processing and data analysis using IR Biotyper® system (Bruker Daltonics, Germany).

WN 1800-900/cm ⇒ carbohydrate + proteins

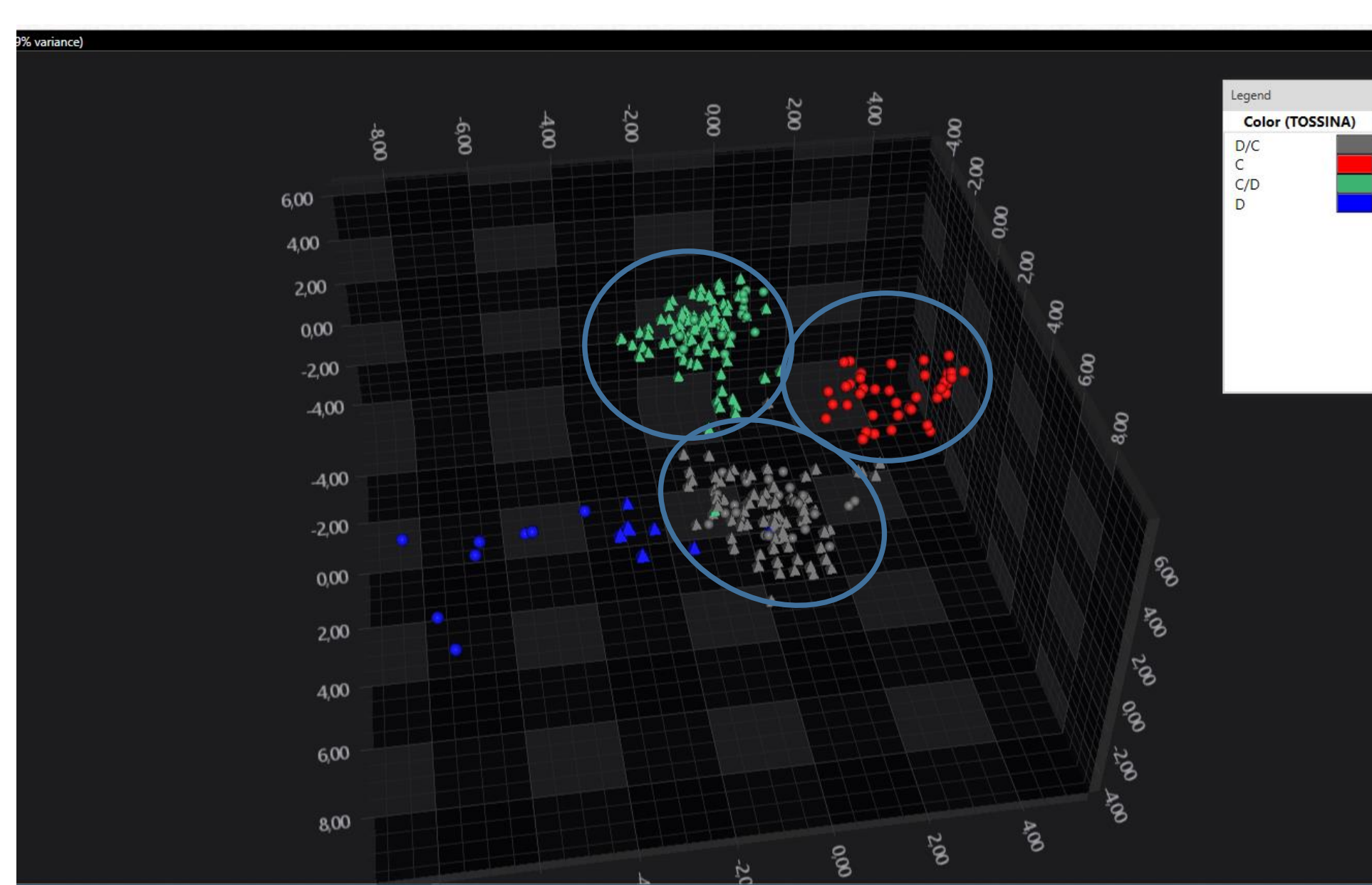


WN 1700-1300/cm ⇒ proteins + lipids



- A, B and F strains clearly differentiable from C, D, CD, DC
- A and B not clearly distinguishable whereas F strain appear clearly different
- C, CD, DC and D strains cluster together with *C. novyi* (Group II *C. botulinum*)

- Three clusters composed exclusively by C, CD and DC strains
- The 2 D strains seems to be quite different: MR 62 non-toxic strains, 4150/24/18 toxic strain.



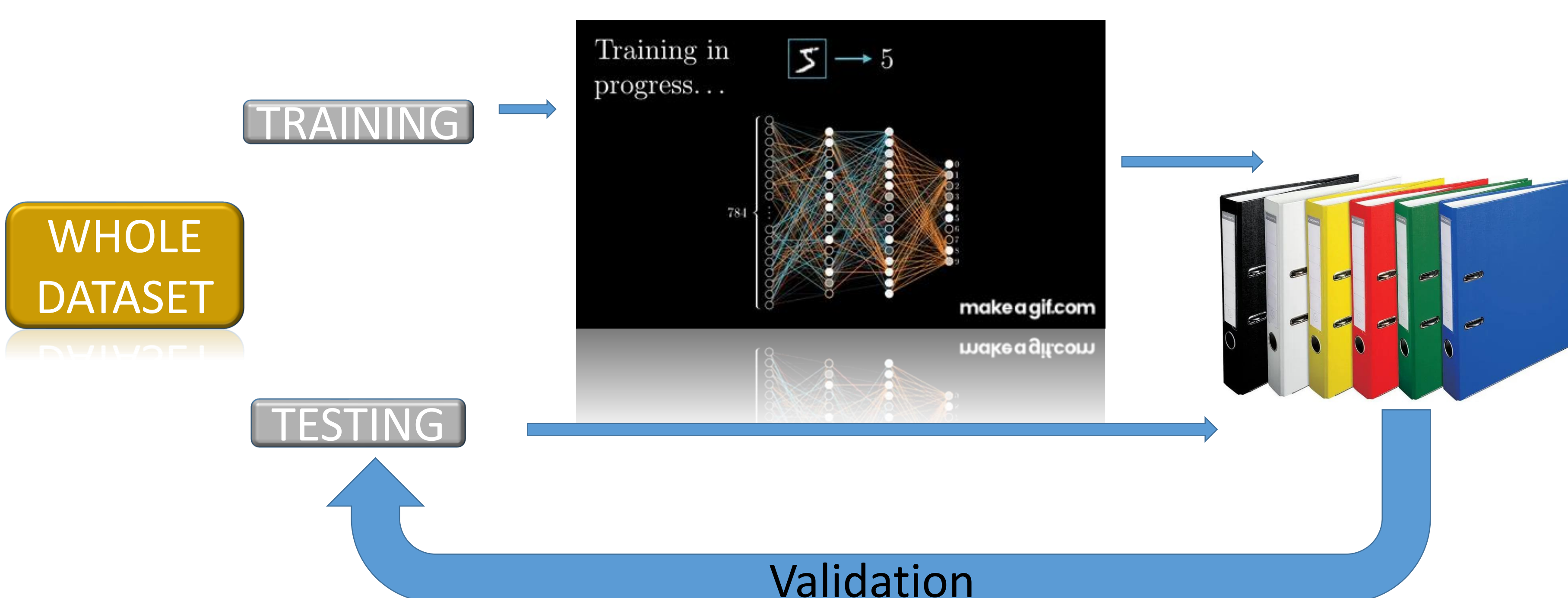
PCA analysis of C, CD, DC and D strains data:

- For *C. botulinum* strains type C, CD, and DC, 3 well-defined clusters are distinguishable in the multidimensional space with high concordance with the BoNT type.
- The two type D isolates look very different in IRBT. Given the limited number of type D strains analyzed so far, it is not yet possible to draw conclusions about the instrument's ability to identify this toxin type as well.

Machine learning algorithm

- Training set: 20 strains representing the biological variability of each group
- Testing set: 12 strains
- Linear Support Vector Machine algorithm (SVM)

Accuracy of SVM: 96% of the testing set



CONCLUSIONS: FT-IRS shows promise as a novel, high-throughput, and routine-friendly method for BPS typing. However, further investigation is needed to solidify these initial findings.