

THE COMPARE DATABASE: A COMPREHENSIVE PUBLIC RESOURCE FOR ALLERGEN IDENTIFICATION AND PROTEIN ALLERGENICITY ASSESSMENT

Main author: Lars K. Poulsen (Gentofte Hospital, University of Copenhagen)

Co-authors: Laurent Beuf, Gabriele Gadermaier, Zhongshan Gao, Eva Gietl, R Karin, Liisa Koski, Justin McDonald, Clare Narrod, Elda Posada-Campos, Andre Silvanovich, Ping Song, Wiebke Striegel, Suzanne Teuber, Ronald van REE, Lucilia Pereira Mouriès

INTRODUCTION

Protein allergenicity is associated with specific protein structure and amino acid sequences that bind immunoglobulin E, as opposed to being linked to protein function. Therefore, a sequence database composed of validated allergens is seen as essential for the safety assessment of genetically modified foods and feeds, as well as other innovative food/feed products. However, the identification of new allergen sequences has become more complex due to the exponential increase of sequence data that accompanied the broad adoption of high throughput sequencing methods.

METHODOLOGY

To address these challenges, an international collaborative scientific group coordinated by the Health and Environmental Sciences Institute (HESI) was tasked to develop a contemporary, adaptable, high-throughput process to build the COMprehensive Protein Allergen RESource (COMPARE) database. The COMPARE database is updated annually using a transparent, computer-based process to identify candidate sequences in scientific literature and publicly accessible protein sequence databases. Candidate sequences are then reviewed and approved by an independent international panel of academic and clinical allergy experts for inclusion in COMPARE.

RESULTS

The COMPARE database, together with its built-in bioinformatic sequence search and visualisation tool known as COMPASS, and extensive documentation, are publicly available at <https://comparedatabase.org/>. The incorporation of the FASTA sequence similarity search algorithm as the core tool in the COMPARE database supports real-time bioinformatic comparisons using global allergen sequence similarity metrics that are based on FAO/WHO and CODEX Alimentarius guidelines. The entire list of sequences and accompanying metadata can be downloaded from the website.

DISCUSSION

This presentation will highlight the most recent database update cycle culminating in the January 2022 COMPARE release. Altogether, the COMPARE database and its publicly available resources support the development of innovative food products and technologies by providing up-to-date tools for allergenicity assessment before new products enter the market, and thus contribute to better protection of human health and food safety.