



DiNovo enables high-coverage and high-confidence *de novo* peptide sequencing via mirror proteases and deep learning



Zixuan Cao¹, Xueli Peng¹, Di Zhang², Piyu Zhou¹, Li Kang³, Hao Chi⁴, Ruitao Wu², Zhiyuan Cheng¹, Yao Zhang³, Jiaying Dai³, Yanchang Li³, Lijin Yao², Xinming Li², Yaoyu He¹, Jinghan Yang¹, Haipeng Wang², Ping Xu³, Yan Fu¹

¹Academy of Mathematics and Systems Science, CAS; ²Shandong University of Technology; ³Beijing Institute of Lifeomics; ⁴Institute of Computing Technology, CAS

Introduction

Despite the recent advancements driven by deep learning, *de novo* peptide sequencing is still constrained by incomplete peptide fragmentation and insufficient protein digestion in current single protease-based proteomic experiments. Here, we present a software system, named DiNovo, for high-coverage and high-confidence *de novo* peptide sequencing by leveraging the complementarity of mirror proteases. DiNovo is empowered by several innovative algorithms, including a mirror-spectra recognition algorithm independent of pre-sequencing, deep neural network-based sequencing algorithm, and target-decoy mapping, a method for sequencing result evaluation free of prior peptide identification. Compared with the trypsin protease used alone, DiNovo using two pairs of mirror proteases leads to two to three times high-confidence amino acids sequenced. Compared with previous single-protease *de novo* sequencing algorithms, DiNovo achieves much higher sequence coverage. DiNovo also shows great potential as a practical and powerful alternative to database search for peptide identification with quality control.

Methods

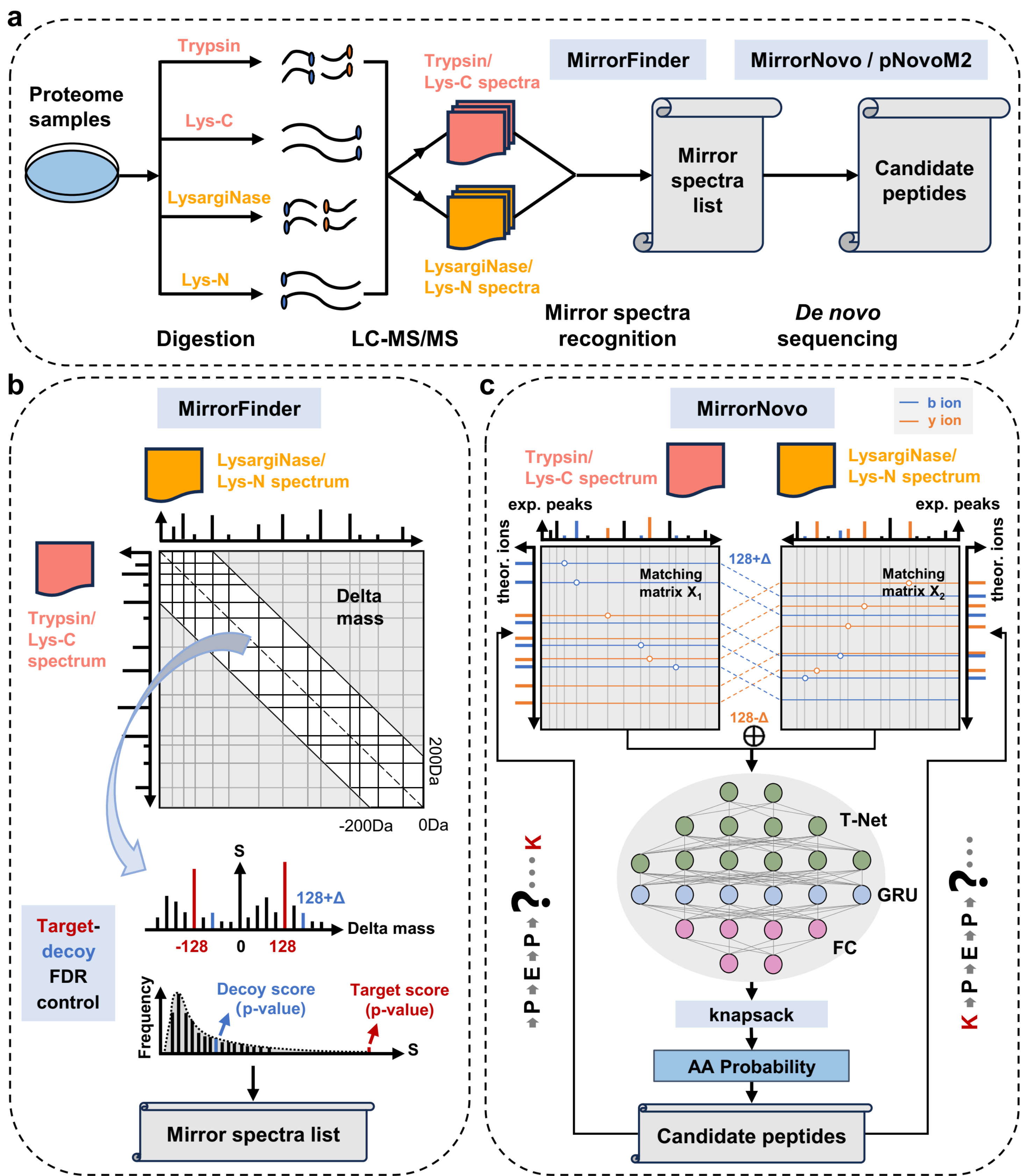


Fig. 1 | Algorithmic workflow and modules of DiNovo

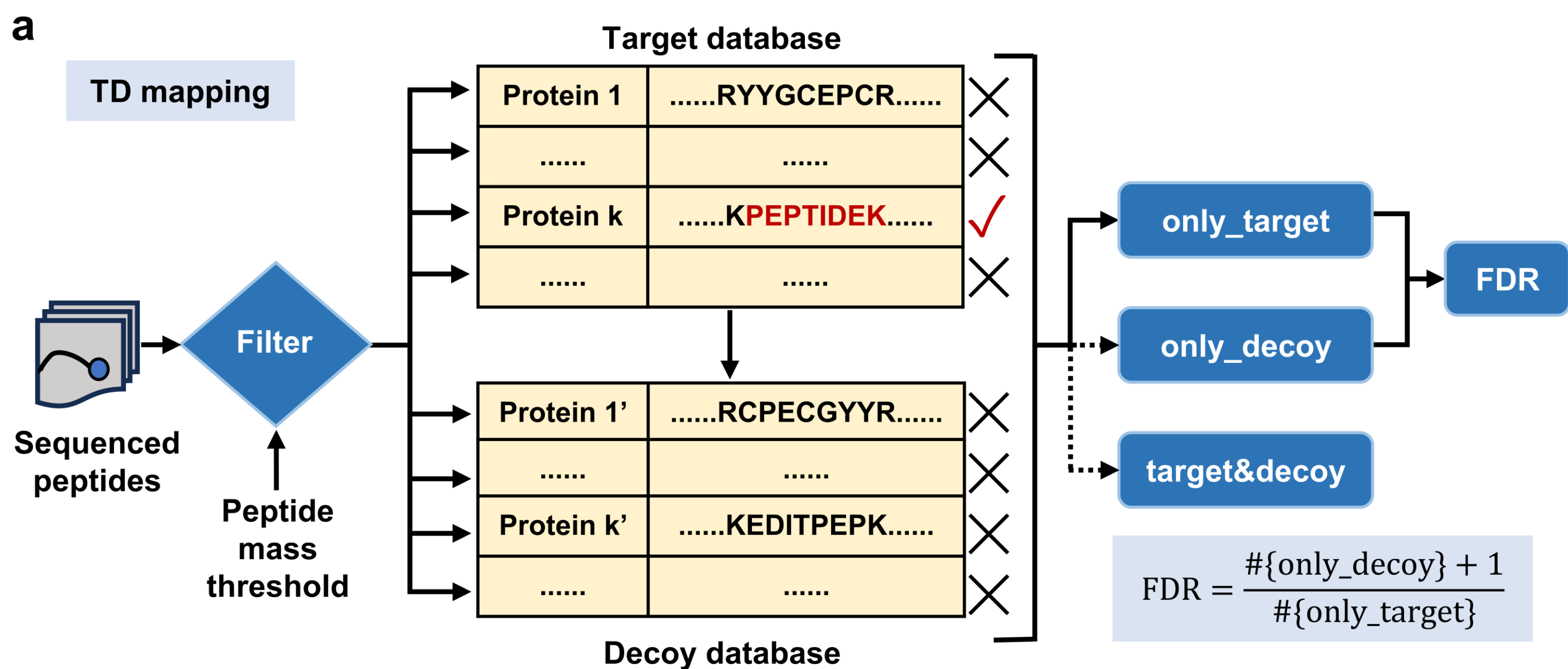


Fig. 2 | Quality control of *de novo* sequencing results

Results

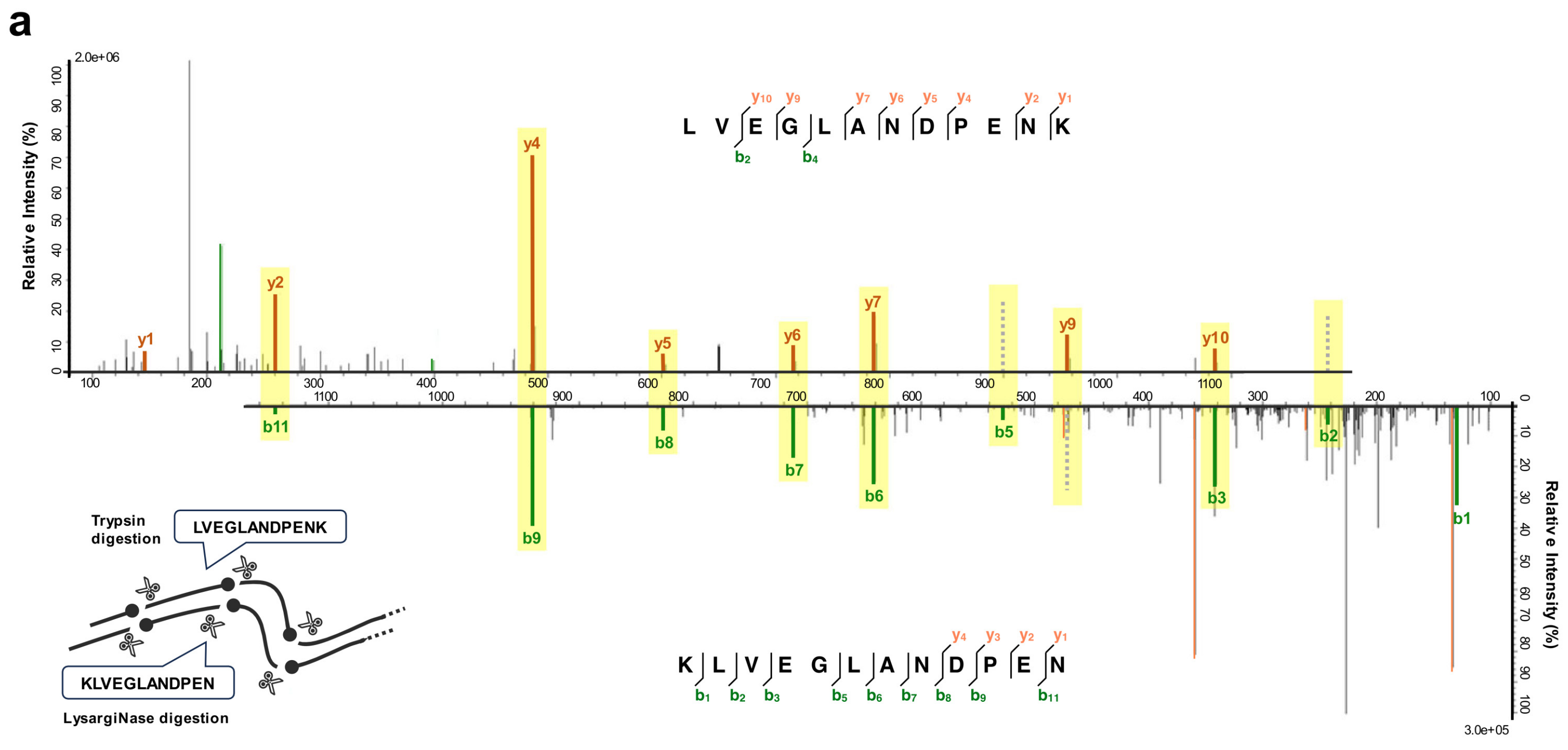


Fig. 3 | Complementarity of mirror spectra

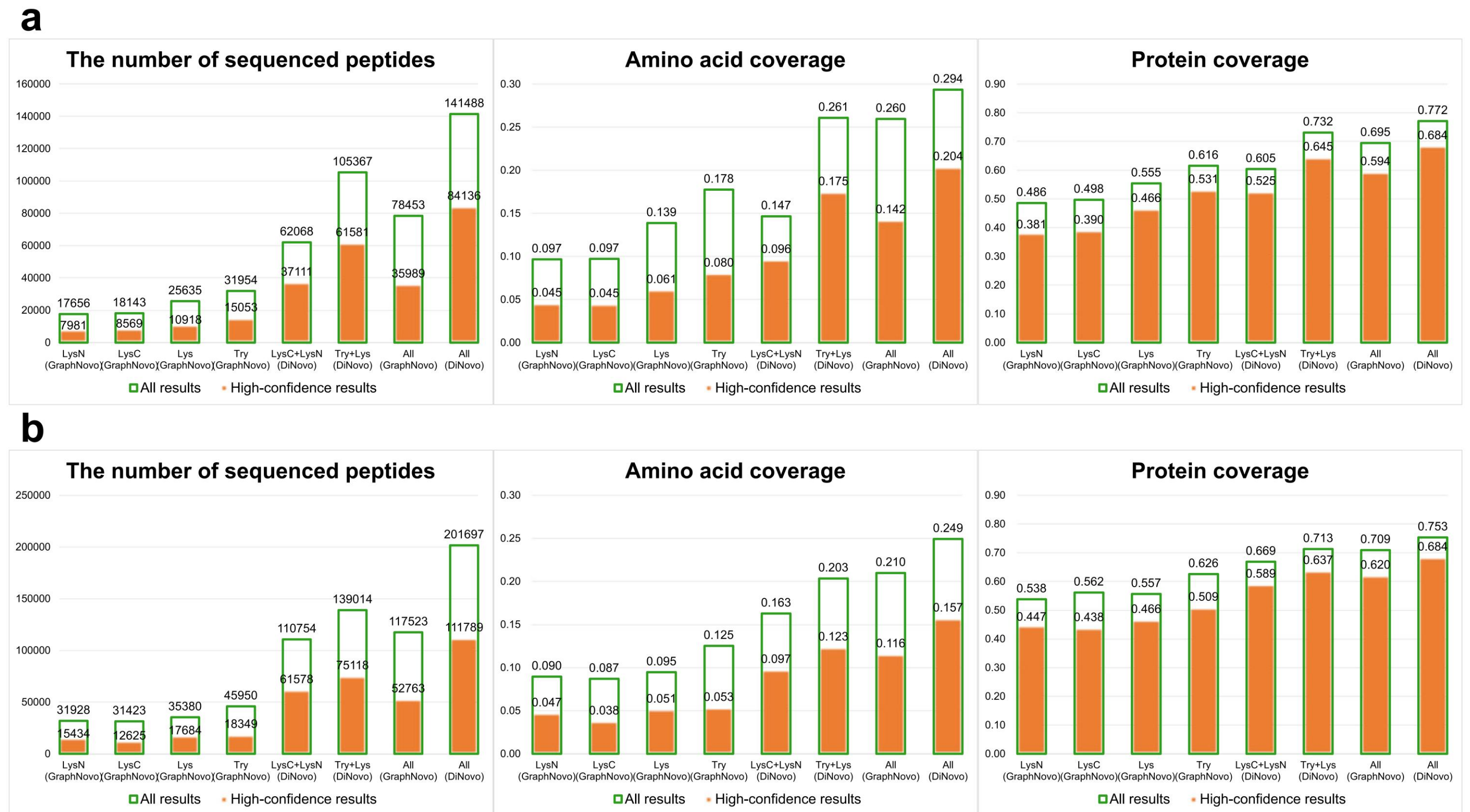


Fig. 4 | Comparison of single- and mirror-protease strategies

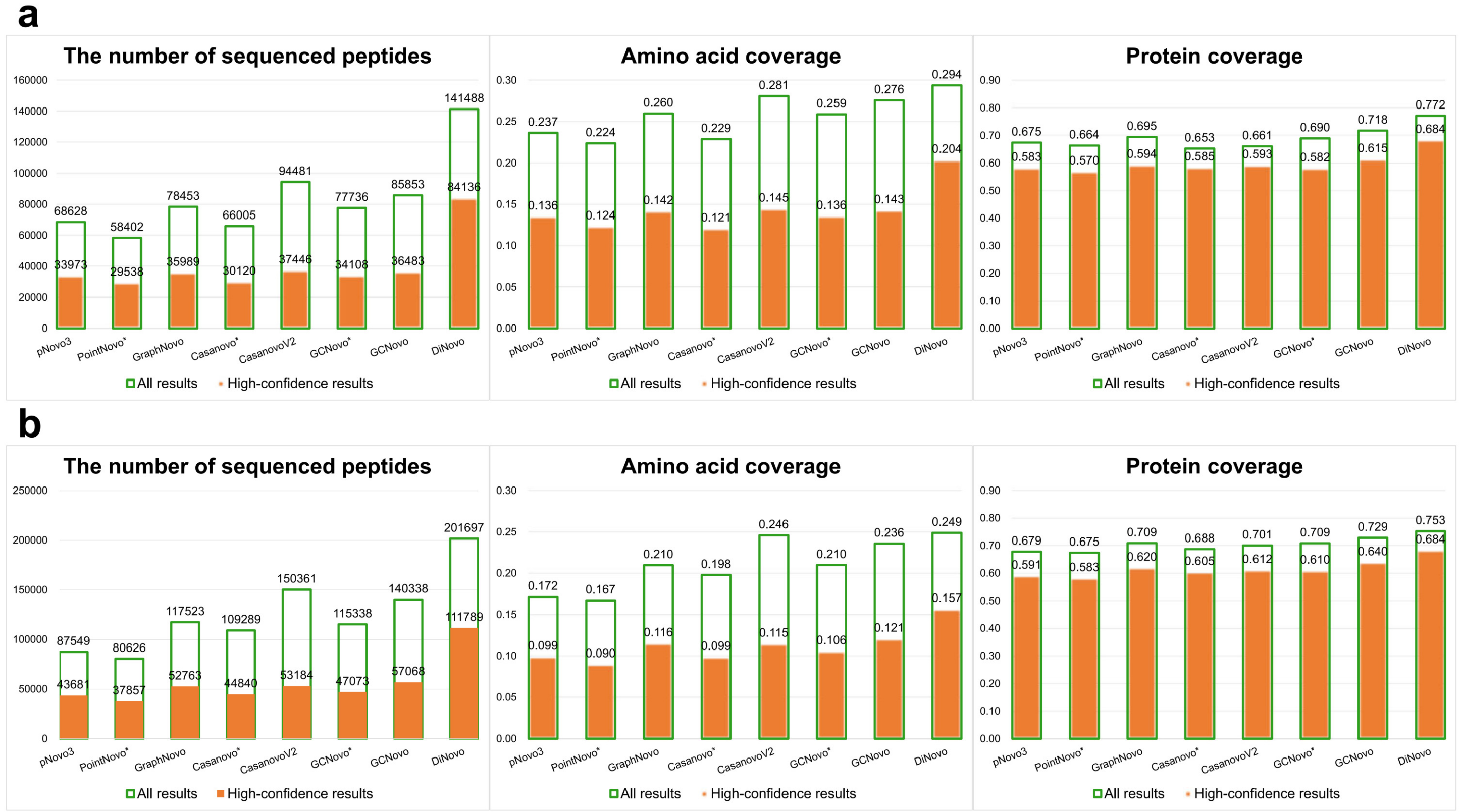


Fig. 5 | Comparison with different *de novo* sequencing algorithms

Reference