

Novel algorithms and the benefits of comparative validation

Robert Smith^{1,*}, Dan Ventura¹ and John T. Prince²

¹Department of Computer Science and ²Department of Chemistry, Brigham Young University, Provo, UT 84602, USA

Associate Editor: Martin Bishop

Contact: 2robsmith@gmail.com

Received on February 22, 2013; revised on April 9, 2013; accepted on April 10, 2013

Bioinformatic research has produced a large volume of proposed algorithmic solutions to a host of problems. Whether presented as a processing step in a clinical experiment or treated in a stand-alone publication, novel bioinformatic algorithms are often not subjected to the thorough comparative evaluation endured by their counterparts in other closely related fields—such as computer science—where an algorithm unevaluated against extant methods is considered unpublishable. Two audiences are interested in algorithmic publications: the practitioner, who may use the algorithm, and the researcher, who will work to develop solutions superior to those extant. We argue that failure during the review/publication process to require comparative evaluation for novel algorithms is detrimental to both parties.

To demonstrate the dilemma, we conducted a case study of novel LC-MS alignment algorithms. Of the 48 publications from 2001 to 2012 that present alignment algorithms of which we are aware, 60% include no comparison with other methods. Another 20% compare their method with one or two others (Fig. 1). Only two articles compare performance against the state-of-the-art methods available at the time of publication. Interestingly, both of these, with six and seven comparisons, respectively, reuse comparative evaluation performance data and datasets from a stand-alone review article of six methods (Lange *et al.*, 2008).

It is natural to wonder whether publication year correlates to the number of comparisons made. After all, earlier articles would have fewer methods to compare against. We found no correlation ($r = 0.397$) between year of publication and number of comparisons (Table 1). Again, the correlation number would be even lower if it was not for the fact that someone published a comparative evaluation of at least some of the extant alignment methods. Without the reuse of that survey article data, the correlation coefficient would drop to 0.313. These data reinforce the prevailing paradigm that comparative performance of a new algorithm to existing ones is too time consuming for the author and reviewers and ought to be the subject of dedicated research (Ballardini *et al.*, 2011). At least for alignment, such dedicated comparison studies are few and far between—we are aware of only one such comparative survey article, even though almost 50 new algorithm articles have been published during the past 11 years (see Lange *et al.*, 2008). Even if these evaluative review articles were more numerous, there are many reasons why

these evaluations ought to be primarily provided in the novel algorithm publications themselves.

A practitioner relies on the peer review process to ensure that the methods they are choosing have met a minimum standard of quality. Though a new method's description or performance may be convincing, these qualities alone are insufficient to weigh the usefulness of an algorithm. Without comparative evaluation, algorithms that underperform against existing ones can easily flood a domain, making the practitioner's task of selecting an algorithm more difficult with every additional publication. Besides an extensive literature review caused by the inundation of articles on the subject, the practitioner must also perform a comparative evaluation of the existing algorithms, as they have no mechanism for quantifying the comparative strengths or weaknesses of the methods from the publications themselves. As pointed out by a recent article, this process is as time consuming as it is difficult, given the oft-encountered difficulties of obtaining and then successfully running someone else's software (Ballardini *et al.*, 2011). Extensive comparative analysis reduces the practitioner's overall time commitment by reducing the number of algorithms under consideration as well as by providing a realistic expectation of performance, hopefully justifying the inevitable inconvenience of obtaining and operating new software. Often, evaluation is made much more difficult (if not impossible) when open source code is omitted in submission. Although English descriptions and pseudocode assist in building intuition about an algorithm, they are lossy definitions that leave out essential details needed for code implementation. Besides time savings, requiring source code facilitates more expansive comparison through automation as well as providing the reviewers an easy metric to determine whether the method is suitably formally defined to be distributed and replicated or whether it is an ad hoc agglomeration.

There are also secondary consequences to consider. Publication is an incentive that can drive innovation. If novel algorithms are not required to outperform extant ones, then innovation—true forward progress not necessarily achieved by mere invention—is less likely to occur. Finding the best choice in an expanding sea of mediocre choices then becomes a Herculean task sure to exhaust any practitioner. The practical result is that practitioners stop short of exhaustively evaluating all the possible options and choose based on some other criteria (e.g. popularity, ease of use or familiarity). The inevitable outcome of the algorithm selection crashshoot are results poorer than what may otherwise have been.

Researchers (the algorithm makers) also suffer when comparative evaluation is neglected. In the face of burgeoning publication numbers, they encounter the same exhaustive search problem faced by the practitioner, but they also face a moral

*To whom correspondence should be addressed.

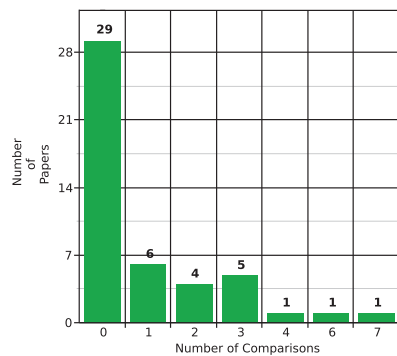


Fig. 1. A comparison of the number of articles presenting MS alignment algorithms and the number of competing algorithms against with they compare. The majority of novel alignment method articles fail to compare against even one extant method

dilemma—the current environment makes it easy to generate many publications, yet difficult to perform the sort of due diligence comparison advocated in this letter. A good comparison requires choosing among the several existing evaluation methods, each of which highlights only specific behavior. The choice is non-trivial—in alignment, metrics include metrics that evaluate the alignment in isolation (Christin *et al.*, 2008, 2010; Van Nederkassel *et al.*, 2006), in combination with other data-processing steps (Ballardini *et al.*, 2011; Lange *et al.*, 2008), globally, and locally. One must also find datasets, which should include sufficient data representative of the different typical performance-affecting real-world characteristics (e.g. complexity of the data, variability of peptide concentration, number of unique and common peptides, extent and form of retention time shift in the data, etc.). What’s more, there is no disincentive provided for publishing work untested against existing methods. Thus, left to their own devices, will the researcher ever behave in a manner that is not in his best interest, though it is in the best interest of the field? Apparently, not very often. Our experience suggests that the pattern we found in alignment algorithms applies to algorithmic approaches in proteomics and metabolomics generally, and it may extend to other bioinformatics subfields where we have less experience.

So what is the solution? The problem, we have found, does not lie in the lack of venue requirements for performance demonstration against state-of-the-art algorithms. Interestingly, many of the articles with zero comparisons came from journals that explicitly require authors to provide quantitative comparison with state-of-the-art methods. Similarly, though an openly available group of standard datasets and metrics as described here would greatly facilitate the evaluations petitioned for, authors in other fields manage to provide comparisons even without standardized metrics or open frameworks for evaluation.

We suggest that greater care be taken by editors and reviewers to require novel algorithmic contributions to contain a reasonable comparative quantitative evaluation with existing methods. New contributions should also include necessary elements to facilitate future comparisons with other algorithms such as source code and parameter setting guidance. Such an effort will inevitably maximize the outcome of practitioner results,

Table 1. A list of articles presenting novel -omics alignment algorithms

Publication	#Comp	Year	Venue
Fraga <i>et al.</i>	0	2001	<i>Anal Chem</i>
Hastings <i>et al.</i>	0	2002	<i>Rapid Com in MS</i>
Bylund <i>et al.</i>	1	2002	<i>J Chrom A</i>
Torgrip <i>et al.</i>	2	2003	<i>J Chemometrics</i>
Åberg <i>et al.</i>	0	2004	<i>J Chemometrics</i>
Lee <i>et al.</i>	0	2004	<i>Anal Chim Acta</i>
Tomasi <i>et al.</i>	0	2004	<i>J Chemometrics</i>
Eilers	0	2004	<i>Anal Chem</i>
Vorst <i>et al.</i>	0	2005	<i>Metabolomics</i>
Pierce <i>et al.</i>	0	2005	<i>Anal Chem</i>
Walczak <i>et al.</i>	4	2005	<i>Chem Intel Lab Sys</i>
Baran <i>et al.</i>	0	2006	<i>BMC Bioinformatics</i>
Smith <i>et al.</i>	0	2006	<i>Anal Chem</i>
Sadygov <i>et al.</i>	0	2006	<i>Anal Chem</i>
Fischer <i>et al.</i>	0	2006	<i>Bioinformatics</i>
Jaitly <i>et al.</i>	0	2006	<i>Anal Chem</i>
Prince <i>et al.</i>	1	2006	<i>Anal Chem</i>
Skov <i>et al.</i>	0	2007	<i>J Chemometrics</i>
Yao <i>et al.</i>	0	2007	<i>J Chrom A</i>
Kirchner <i>et al.</i>	0	2007	<i>J Stat Software</i>
Palmblad <i>et al.</i>	0	2007	<i>ASMS</i>
Lange <i>et al.</i>	0	2007	<i>Bioinformatics</i>
Wang <i>et al.</i>	0	2007	<i>Biostatistics</i>
Mueller <i>et al.</i>	0	2007	<i>Proteomics</i>
Listgarten <i>et al.</i>	0	2007	<i>Bioinformatics</i>
Fischer <i>et al.</i>	2	2007	<i>BMC Bioinformatics</i>
Csenki <i>et al.</i>	3	2007	<i>Anal Bioanal Chem</i>
Åberg <i>et al.</i>	0	2008	<i>J Chrom A</i>
De Groot <i>et al.</i>	0	2008	<i>Proteomics</i>
Suits <i>et al.</i>	0	2008	<i>Anal Chem</i>
Shinoda <i>et al.</i>	0	2008	<i>Bioinformatics</i>
Christin <i>et al.</i>	1	2008	<i>Anal Chem</i>
Podwojski <i>et al.</i>	2	2009	<i>Bioinformatics</i>
Befekadu <i>et al.</i>	3	2009	<i>IEEE EMBS</i>
Christin <i>et al.</i>	3	2010	<i>JPR</i>
Daszykowski <i>et al.</i>	0	2010	<i>J Chrom A</i>
Tomasi <i>et al.</i>	1	2010	<i>J Chrom A</i>
Bloemberg <i>et al.</i>	1	2010	<i>Chem Intel Lab Sys</i>
Eliasson <i>et al.</i>	0	2011	<i>Curr Pharm Biotech</i>
Sinkov <i>et al.</i>	0	2011	<i>Anal Chim Acta</i>
Befekadu <i>et al.</i>	3	2011	<i>IEEE ACM TCBB</i>
Tang <i>et al.</i>	3	2011	<i>Prot Science</i>
Ballardini <i>et al.</i>	6	2011	<i>J Chrom A</i>
Voss <i>et al.</i>	7	2011	<i>Bioinformatics</i>
Zhang	0	2012	<i>ASMS</i>
Struck <i>et al.</i>	1	2012	<i>J Chrom A</i>
Hoekman <i>et al.</i>	2	2012	<i>ASBMB</i>
Kaya <i>et al.</i>	3	2012	<i>Inform Sciences</i>

Note: The data have a correlation coefficient of 0.397, suggesting there is no trend toward comparison against extant algorithms.

encourage the widespread use of the highest-quality tools and provide researchers an incentive to truly innovate.

Funding: National Science Foundation Graduate Research Fellowship [DGE-0750759].

Conflict of Interest: none declared.

REFERENCES

- Ballardini,R. *et al.* (2011) MassUntangler: a novel alignment tool for label-free liquid chromatography–mass spectrometry proteomic data. *J. Chromatogr. A*, **1218**, 8859–8868.
- Christin,C. *et al.* (2008) Optimized time alignment algorithm for LC-MS data: correlation optimized warping using component detection algorithm-selected mass chromatograms. *Anal. Chem.*, **80**, 7012–7021.
- Christin,C. *et al.* (2010) Time alignment algorithms based on selected mass traces for complex LC-MS data. *J. Proteome Res.*, **9**, 1483–1495.
- Lange,E. *et al.* (2008) Critical assessment of alignment procedures for LC-MS proteomics and metabolomics measurements. *BMC Bioinformatics*, **9**, 375.
- Van Nederkassel,A. *et al.* (2006) A comparison of three algorithms for chromatograms alignment. *J. Chromatogr. A*, **1118**, 199–210.